Reading and Writing (2024) 37:385–417
https://doi.org/10.1007/s11145-022-10331-0

Global temporal typing patterns in foreign language writing: exploring language proficiency through recurrence quantification analysis (RQA)

Lisa Haake[1] · Sebastian Wallot[2,3] · Monika Tschense[2,3] · Joachim Grabowski[1]

Accepted: 2 July 2022 / Published online: 8 August 2022
© The Author(s) 2022

Abstract
Recurrence quantification analysis (RQA) is a time-series analysis method that uses autocorrelation properties of typing data to detect regularities within the writing process. The following paper first gives a detailed introduction to RQA and its application to time series data. We then apply RQA to keystroke logging data of first and foreign language writing to illustrate how outcome measures of RQA can be understood as skill-driven constraints on keyboard typing performance. Forty native German students performed two prompted writing assignments, one in German and one in English, a standardized copy task, and a standardized English placement test. We assumed more fluent and skilled writing to reveal more structured typing time series patterns. Accordingly, we expected writing in a well-mastered first language to coincide with higher values in relevant RQA measures as compared to writing in a foreign language. Results of mixed model ANOVAs confirmed our hypothesis. We further observed that RQA measures tend to be higher, thus indicating more structured data, whenever parameters of pause, burst, and revision analyses indicate more fluent writing. Multiple regression analyses revealed that, in addition to typing skills, language proficiency significantly predicts outcomes of RQA. Thus, the present data emphasize RQA being a valuable resource for studying time series data that yields meaningful information about the effort a writer must exert during text production.

Keywords Recurrence quantification analysis · Time course data · Foreign writing research · Language proficiency · Writing fluency

Writing in a first language (L1) can be challenging as authors need to manage the simultaneous engagement of various tasks that compete for attention. Skilled writers

✉ Lisa Haake
haake@psychologie.uni-hannover.de

Extended author information available on the last page of the article

know how a text should be composed regarding structure, content, and tone. They know linguistic forms that allow the nuanced expression of ideas and can efficiently coordinate the different tasks that are involved during writing (Cumming, 1989). Writing skills can be used across languages: Strategies that are applied in L1 can be transferred when writing in a foreign language (Riehl, 2014). Yet, writing in a foreign language is usually more effortful and the individual level of language proficiency affects the fluency with which a text is compiled (Van Waes & Leijten, 2015). As writing processes are coordinated in time and their products are characterized by a sequential structure, analyzing time-series data from typing activity on a keyboard is an efficient method to reveal underlying motor-cognitive processes during writing (Galbraith & Baaijen, 2019). Analyzing keystroke logging data might thus be particularly powerful to investigate how language proficiency shapes writing processes.

Recurrence quantification analysis (RQA; Webber & Zbilut, 1994) is a time-series analysis method that uses keystroke logging data to recruit information from autocorrelation properties of the typing data (Wallot & Grabowski, 2019). Autocorrelation is a means to determine how data points collected over time form clusters or tend to occur repeatedly in terms of fluctuation patterns. Other than, for example, the Pearson correlation which quantifies the association of two variables in terms of magnitude, direction, and R^2, RQA connects data points across time series of keystrokes. By quantifying the dynamics of writing via such global fluctuation patterns, RQA can yield valuable information about the writing process in its entirety.

Conventional approaches that use keystroke logging data to assess writing processes typically analyze a wide range of parameters such as the locations and times a writer paused (i.e., the interruption of typing activity), numbers and lengths of burst productions (i.e., the continuous string of characters typed between two pauses or revisions), or revisions (i.e., adding sequences to or deleting parts from already produced text). Together, these complementary measures are then used as indicators for the fluency with which a text was composed and, in return, unveil time and location of potentially difficult episodes of text production.

RQA, in contrast, shifts the analysis from local to global typing patterns and hence takes on a more holistic approach: Rather than integrating different aspects of time, location and/or text length, RQA uses autocorrelation properties of the typing data to detect regularities within the writing process. The information it provides can then be diagnostic, so the idea, for the effort a writer must exert during text production and, in return, the general constraints a writer had during the composition.

In the perspective of dynamic systems analysis, "constraints" are a controlling factor for the dynamics of a system (for an introduction to the principles of dynamic systems, see Kelso, 1995). Generally speaking, constraints are factors on time scales slower than the actual behavior of the system that delimit the space in which the system dynamics evolve. As a result, constraints disallow certain dynamics that a system is principally capable of exhibiting. In terms of writing (or reading), writing skills control performance. The dynamic systems analogy to constraints would imply that writing skill evolves slower than the performance in a particular writing task at hand, reducing the different patterns of finger movements on a keyboard that two hands with five fingers each could principally exhibit.

Admittedly, "skill" in writing is not a single, homogenous category, but a generic one that comprises different facets like language or motor control ability, and it is still an ongoing topic of research of how skill controls writing performance (e.g., Van Waes et al., 2021). However, there are various indicators that writing skill effectively delimits writing performance. A few-finger typing system, for instance, leads to some fingers operating numerous keys. In contrast, skilled 10-finger touch typing leads to close coupling of each finger with a much more limited set of keys that greatly reduces search movements (e.g., van Weerdenburg et al., 2019). Accordingly, unskilled writers show larger typing intervals due to finger movement trajectories when searching for the right key on a keyboard (Jokinen et al., 2017). Further, training of manual motor coordination (Smethurst & Carson, 2001) or tapping (Riley et al., 2012) leads to more stable and confined dynamics of movement. Following from that, typing on a keyboard to write in a well-known L1 should be different from typing on a keyboard to write in English as a foreign language (EFL).

As we will explain in detail below, RQA captures fluctuation and autocorrelation properties of a time series. Thereby, RQA yields several outcome measures that relate to the dynamic stability of that time series, indicating whether the dynamics (of typing intervals) are more structured or rather unstructured over time. These outcome measures can then be interpreted in terms of the degree of constraints put on the writing process. Language proficiency is a constraint that, amongst others, delimits writing performance. By manipulating a writing task (writing in L1 or EFL), and using language proficiency (derived from a language aptitude test) and typing skill (assessed via a standardized copy task) information, we aim to illustrate how RQA outcome measures can be understood as skill-driven constraints on keyboard typing performance.

Studies investigating the challenges writers face when writing in a foreign language showed that production fluency tends to be lower when writing in a foreign language as opposed to a well-known L1: Writers are more likely to produce shorter texts and pause more frequently (Van Waes & Leijten, 2015), and the number of characters written between two pauses or revisions tends to be significantly lower (Chenoweth & Hayes, 2001, 2003; Hayes & Chenoweth, 2006). Previous studies investigating reading fluency using RQA revealed that reading times with higher dynamical structures indicated more fluent, integrated, and thus proficient reading (O'Brien & Wallot, 2016; Wallot et al., 2014). Furthermore, a study on copy-typing was the first to show conceptually similar effects for typewriting (Wallot & Grabowski, 2019). Accordingly, we expect to find more fluent and skilled writing to reveal more structured typing time series patterns. Thus, writing in a well-mastered L1 should coincide with higher values in relevant RQA measures in contrast to writing in EFL. Further, typing patterns should reveal more regularities whenever conventional variables exploring pauses, bursts, or revisions indicate more fluent writing processes.

The paper is structured as follows: We will first introduce RQA and its application to time-series data. We will illustrate how recurrences and parameters are defined within the RQA-framework, how parameters can be estimated for sample data, and define most important outcome measures to show how patterns of keystroke activity are reflected in them. Subsequently, we will apply RQA to actual

keystroke logging data of first and foreign language writing. We will compare outcomes of RQA to more conventional approaches of fluent writing measurements such as parameters of pausing, burst productions, and revising. To describe RQA from a multi-dimensional perspective and scrutinize RQA's potential in foreign language writing contexts, we will then examine how also gradual changes of language proficiency affect outcomes of RQA. Thereby, we aim to demonstrate that RQA is a valuable resource for determining the general constraints a writer has during writing that provides meaningful data and enriches the methodological inventory for analyzing time course data.

Recurrence quantification analysis (RQA)

In this section, we are going to provide a detailed description of RQA. As the name already implies, RQA is about repetitions of sorts. The core of RQA is the recurrence plot, which charts recurring patterns of values in a time series or sequence (Eckmann et al., 1987). In effect, the recurrence plot is a two-dimensional depiction of a one-dimensional series of measurements. Consider the first two lines of the nursery rhyme 'Humpty Dumpty':

Humpty Dumpty sat on a wall,
Humpty Dumpty had a great fall.

We can think of this rhyme as a sequence of characters. Further, it is a sequence of nominal values, in which values are either identical ($A = A$) or not ($A \neq B$). Recurring instances of identical values are depicted in a recurrence plot, as shown in Fig. 1 for the 'Humpty Dumpty' rhyme. The recurrence plot is a binary matrix that results from the cross-comparison of the values of a time series. The values of the matrix are coded as 1 when two values are identical and hence counted as recurrent, indicated as black dots in the matrix. Non-identical values are coded as 0, which means they are non-recurrent, making up the white space of the matrix.

There are a couple of features to note within the recurrence plot: First, there is an uninterrupted line of recurrence running along the main diagonal, from the lower-left to the upper-right. This line of identity is always present in a recurrence plot, even for randomly drawn numbers. The line of identity charts recurrences at lag 0 by comparing every value in the sequence with itself at the same time and position. This also means that in the case of an auto-recurrence plot the line of identity does not give information about the dynamics of a time series, or structure of a sequence, and should be discarded. Second, the recurrence plot is symmetrical above and below the line of identity such as the recurrence patterns in the lower-right half of the plot are mirror versions of the recurrence patterns in the upper-left half of the plot, both capturing lagged recurrence patterns in the sequence. When we inspect diagonals to the right and left of the line of identity, it is akin to shifting our sequence by a certain number of letters and examining how recurrences change consequently, resulting in lagged patterns. The first diagonal to the left or right of the line of identity charts for example auto-recurrences at lag 1, i.e., when we move the sequence by one letter forward or backward. Finally, the recurrences are not evenly

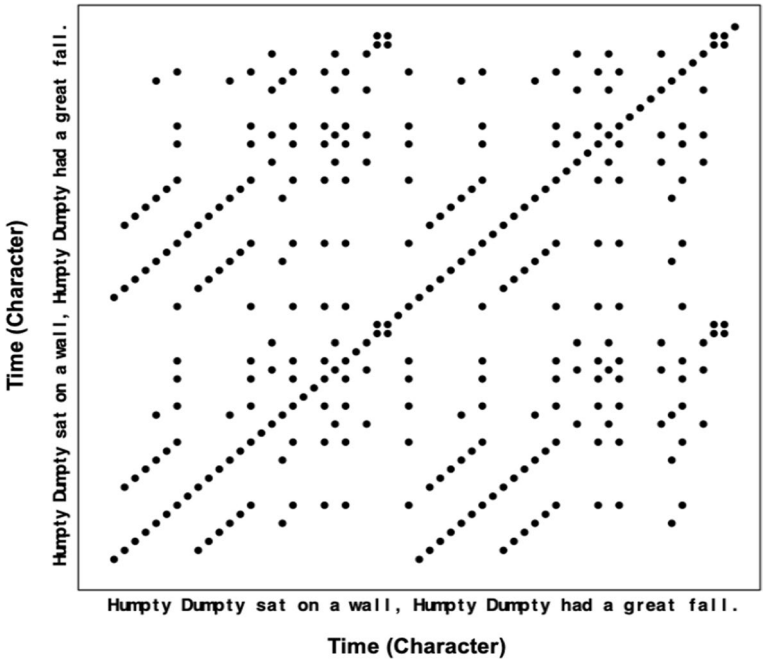

Fig. 1 Recurrence plot for the first two lines of the 'Humpty Dumpty' nursery rhyme. Black dots are recurrence points that indicate where characters are repeated in the sequence

distributed across the plot. Rather, we see that certain areas of the plot are more densely populated than others and that recurrences cluster differently across the plot. This clustering captures the structure within our 'Humpty Dumpty' example, e.g., its rhyming patterns.

Nominal sequences, such as a series of characters, illustrate the concept of recurrence well. However, when analyzing keystroke logging data, we usually have interval- or absolute-scale data. Imagine, we had recorded the inter-key intervals (IKIs) of a participant writing these first two lines of 'Humpty Dumpty'. To provide an example, we have fabricated a putative time series of IKIs that could have created this text as illustrated in the top panel of Fig. 2. Here, the hypothetical participant had a good idea of how to write the first line, typing it quickly in a series of strokes while being less certain about the second line, resulting in more interspersed fast and slow IKIs. Further, we suppose that the participant had an idea of how to change and finish line two which results in rapidly erasing some characters, and then quickly finishing this line in a sequence of swift strokes.

If we insist on the identity of values to define recurrences, the corresponding recurrence plot will be empty apart from the line of identity (and in our example also the key delay of 60 ms, when some of the characters were successively erased very quickly; Fig. 2—bottom left), because two values do not occur strictly equal. As can be seen in the lower-left panel of Fig. 2, such a recurrence plot is not very informative. The only thing it tells us about the dynamics of the IKIs is that there is

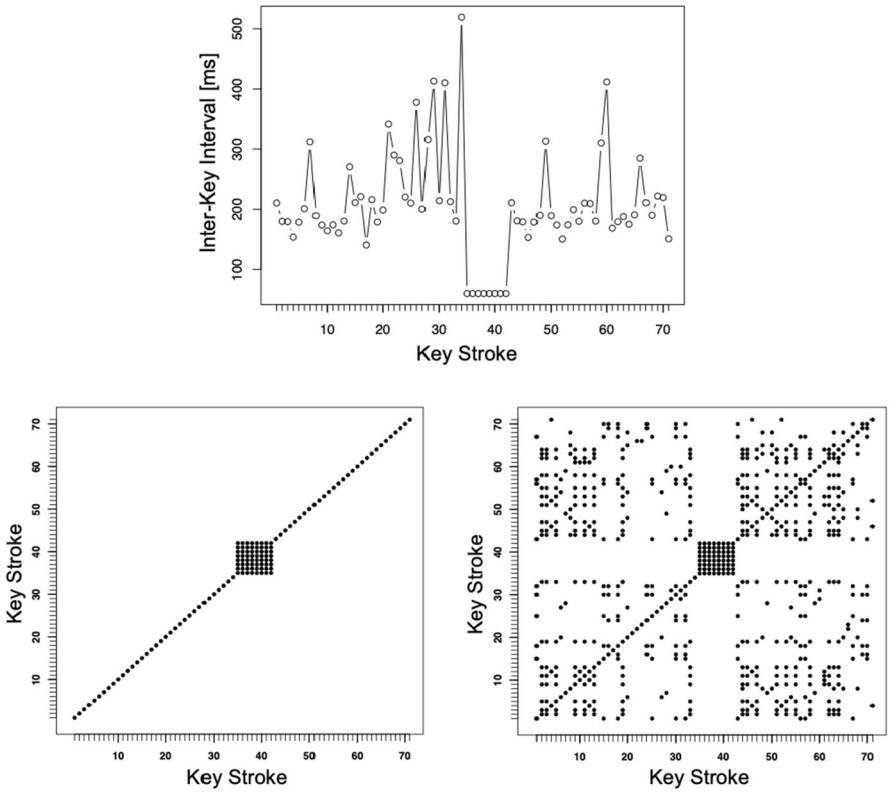

Fig. 2 Top: putative time series of inter-key intervals (IKIs; given in ms) for typing the first two lines of the 'Humpty Dumpty' nursery rhyme. Bottom: corresponding recurrence plots with $r=0$ (left) and $r=0.1$ (right)

one single patch of identical values that results from very fast, repetitive key presses. To visualize and then quantify the dynamics of such a time series, we need to find (approximate) recurrences that reveal further structure in the data. To do so, we need to apply a threshold parameter r, which provides a range around each value, so that we count similar, but not identical values as recurrent. While nominal sequences are calculated with a threshold parameter of $r=0$ so that only identical values are counted as recurrent, we need to apply some threshold parameter of $r>0$ to find recurrences when using continuously measured data. Such a recurrence plot is shown in Fig. 2—bottom right.

A formal definition of recurrence plots and RQA measures

Now that we are familiar with recurrence plots, let us turn to their formal definition. The basis of each recurrence plot is a distance matrix, which charts the distance between each data point in a time series:

$$DM_{i,j} = \left| x_i - x_j \right|, \quad i,j = 1, \ldots, N \tag{1}$$

where x corresponds to the time series of data, N is the length of the time series (i.e., the number of data points of x), and | | indicates the absolute value. This distance matrix is then thresholded by a threshold parameter r to define recurrent and non-recurrent points:

$$RP = \Theta(r - \|DM\|) \tag{2}$$

where Θ represents the Heavyside step function, r is the threshold parameter, || || indicates a norm, and DM corresponds to the distance matrix.

The resulting thresholded distance matrix is now a binary matrix where recurrences (i.e., distances between two data points $\leq r$) are coded as 1, and non-recurrences (i.e., distances between two data points $> r$) are coded as 0. This binary matrix is what we have displayed as recurrence plots in Figs. 1 and 2 which also provides the basis for calculating recurrence measures.

While the recurrence plot as such features a qualitative, visual display of the dynamics of a time series, we are usually interested in quantifying the properties of a time series. Three kinds of patterns are of particular importance for most measures that can be defined: isolated recurrence points, diagonally adjacent recurrence points, and horizontally/vertically adjacent recurrence points (Coco & Dale, 2014; Marwan et al., 2007). Figure 3 illustrates these patterns for our putative IKI series.

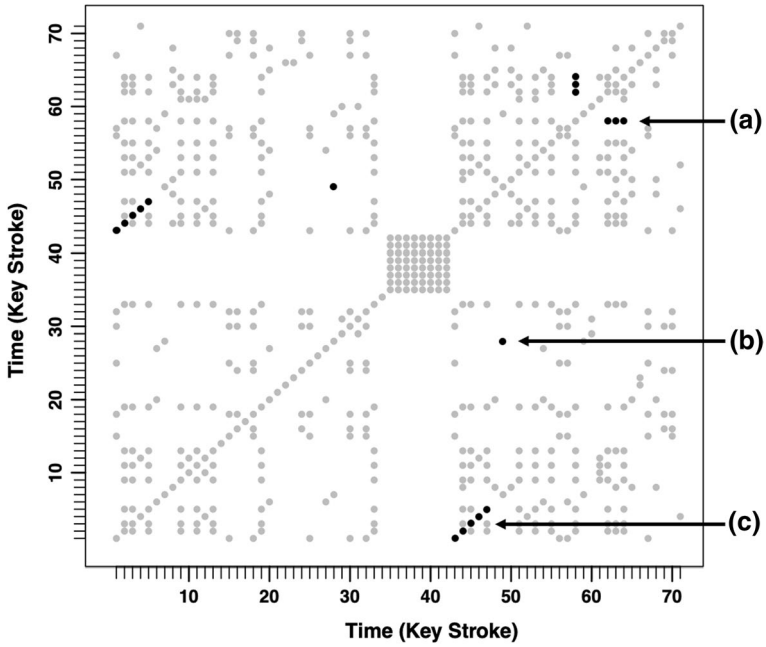

Fig. 3 Illustration of three crucial recurrence patterns: **a** vertically/horizontally adjacent recurrence points, **b** isolated recurrence points, and **c** diagonally adjacent recurrence points

While horizontally/vertically adjacent recurrence points indicate that the time series has settled into a stable (or only slowly changing) state, isolated recurrence points indicate a single match between values. Finally, diagonally adjacent recurrence points show that a whole trajectory of values is repeated in a sequence.

These types of clusters can be quantified by several measures that capture different aspects of dynamics or sequential structure. The simplest measure is the percentage of recurrence points in a plot, known as the recurrence rate. It is the sum of all recurrence points (excluding the line of identity) divided by the sum of all possible recurrence points given by the size of the respective recurrence plot. The recurrence rate mainly captures the occurrence of individual matches of values in a sequence. A measure that captures the structure of the time series in terms of larger patterns is called determinism rate or percent determinism. The determinism rate describes the sum of all recurrence points with diagonally adjacent neighbors divided by the sum of all recurrence points. Thereby, it captures how strongly a time series or sequence is structured in terms of larger sub-trajectories. For the 'Humpty Dumpty' example, the determinism rate shows how many of the individual letters recur as substrings—syllables, words, or multi-word groups. A related measure, the average diagonal line length, quantifies the average length of such sub-strings. There are further recurrence measures that capture different aspects of the dynamics in a time series. Table 1 summarizes some important ones and the three measures briefly presented above. All measures of RQA capture principally independent aspects of the dynamics of a time series. For very stochastic time series, these measures are often strongly correlated among each other and are all more general indicators of the strength of patterning and auto-correlation within a time series (Wallot, 2017). However, for trace measures of handwriting exhibiting a much stronger deterministic component, different measures might indicate quite distinct writing properties.

We can now apply these parameters to the two kinds of data that we have presented so far: nominal sequences and IKIs. Figure 4 provides recurrence plots for both, the original version and a shuffled version of our examples. The respective recurrence measures are shown in Table 2. As can be seen, shuffling a time series does not result in different recurrence rates since data points are still the same. Hence, it is the repetitiveness of a sequence. However, randomizing the order in which data points occur removes its temporal structure, leading to lower values for the determinism rate and the average diagonal line length, as well as most of the other measures.

Parameters and parameter selection

To perform RQA, three parameters must be set: The (time) delay parameter t, the embedding dimension parameter m, and the threshold parameter r. The parameters (time) delay parameter t and the embedding dimension parameter m are necessary to properly unfold the multidimensional dynamics of a time series, allowing us to reconstruct higher-dimensional dynamics—the phase-space of the actual dynamics from which our one-dimensional time series originated (Marwan et al., 2007). There are different options to do so (Lekscha & Donner, 2018). In this paper, we

Table 1 Common recurrence measures

Measure	Definition	Quantifies…
Recurrence rate	Sum of recurrent points divided by the size of the recurrence plot	…the repetition of elements of the time series
Determinism rate	Sum of diagonally adjacent recurrent points (minimal length of 2, LOI excluded) divided by the sum of recurrent points	…the number of individual repetitions occurring in connected trajectories
Average diagonal line length	Average length of diagonal lines (minimal length of 2, LOI excluded)	…the average length of connected trajectories
Maximum diagonal line length	Length of longest diagonal line (LOI excluded)	…the longest connected trajectory
Entropy	Entropy of the distribution of diagonal line lengths (minimal length of 2, LOI excluded)	…the transitions between states (bifurcations)
Laminarity	Sum of vertically adjacent recurrent points (minimal length of 2, LOI excluded) divided by sum of recurrent points	…the number of individual repetitions occurring in laminar states*
Trapping time	Average length of vertical lines (minimal length of 2, LOI excluded)	…the average length of a laminar state*

This table is not an exhaustive list, but rather summarizes commonly used recurrence measures. However, more recurrence measures exist, and further ones are being developed

LOI line of identity

*A laminar state refers to adjacent data points in the time series of constant values or only slowly changing values (e.g., the patch of recurrences in the lower-left panel of Fig. 2 where the same value occurs eight times in succession—here, due to fast, consecutive pressing of a key at sampling rate). See, for example, Coco and Dale (2014), Marwan et al. (2007) and Wallot and Leonardi (2018)

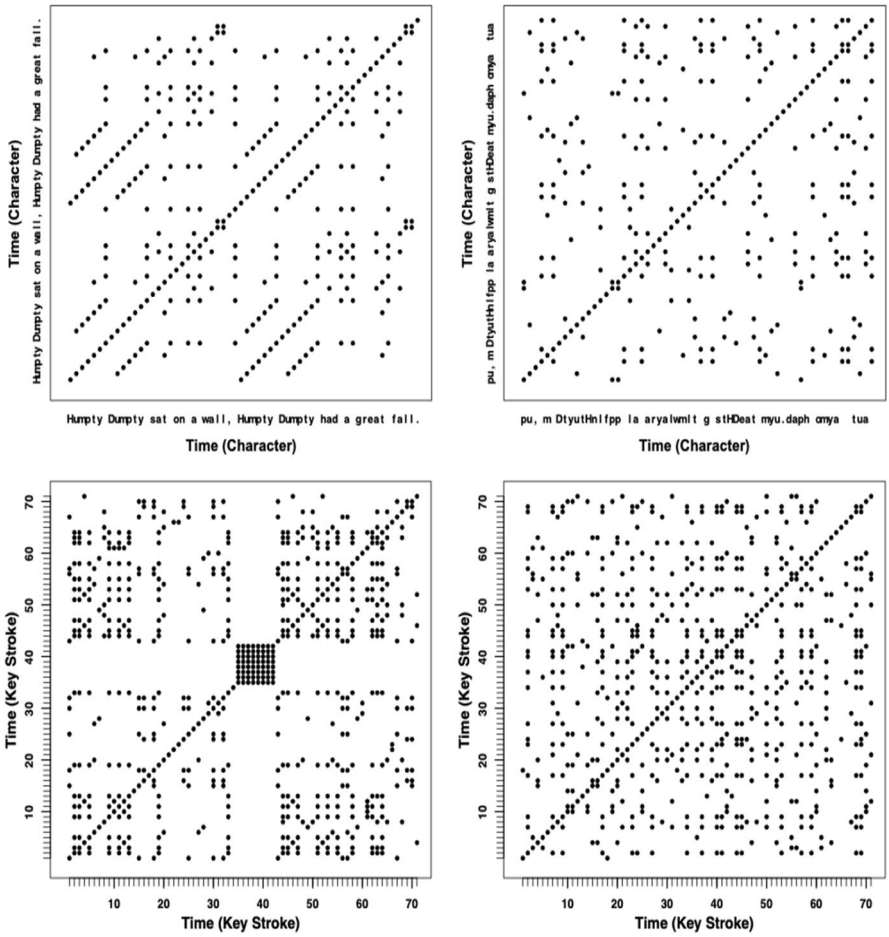

Fig. 4 Recurrence plots for the 'Humpty Dumpty' text (top panel), and the putative inter-key interval data (bottom panel). The left panel refers to the original sequences, while the right panel shows the respective time series in randomized order

will focus on the most used procedure, the time delay embedding method (Takens, 1981) which estimates the delay parameter via average mutual information and the embedding parameter via false nearest neighbors.

Parameter estimations for recurrence-based analyses usually start with the delay parameter t, even though the embedding dimension parameter m would logically have to be estimated first. Why? If the embedding parameter m were to be estimated as $m = 1$, then the time series would not be embedded, and the delay parameter t would not be applied after all. However, to obtain a good estimate for embedding dimensions, the time delay needs to be specified (Wallot, 2017). To estimate the delay parameter t, we use the average mutual information function. The function shows at which lag (or lags) the time series x is least correlated with itself. We want

Table 2 Recurrence measures for example recurrence plots

		RR	DET	ADL	MDL	TT	LAM
Text	Ordered	6.83	38.21	5.88	14.00	2.00	3.25
	Randomized	6.83	14.63	2.00	2.00	2.00	8.94
IKIs	Ordered	11.62	37.54	2.62	7.00	2.49	34.47
	Randomized	11.62	23.55	2.16	4.00	2.01	23.38

Recurrence measures for the character sequence (Text) and the putative series of inter-key intervals (IKIs) for the 'Humpty Dumpty' example

RR recurrence rate, *DET* determinism rate, *ADL* average diagonal line length, *MDL* maximum diagonal line length, *TT* trapping time, *LAM* laminarity

to reconstruct the contributions of other, unobserved variables to the potentially multidimensional dynamics of the system from which x was measured. As explained above, we do this by using time-shifted copies of x, which are now surrogates for the other, unobserved dimensions of the original dynamics. These surrogate series should be as independent as possible from the original series to provide maximum information about the original, multidimensional dynamics. That is, they should contribute as much new, additional information as possible. A means to define 'additional information' is the degree of independence as given by the average mutual information value for a particular lag: the lower the average mutual information, the more independent is the shifted time series, and the more additional information is contributed by a surrogate series. Figure 5 shows the computation of the average mutual information function for the first 20 lags of our putative IKI data. To decide which delay to use, it is common practice to choose the first local minimum of the average mutual information function (Fraser & Swiney, 1986). Since each surrogate

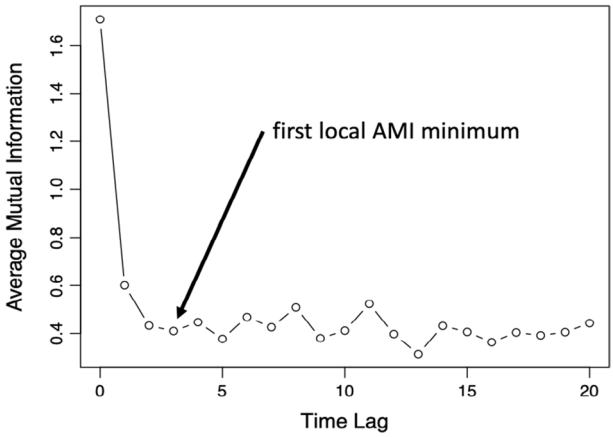

Fig. 5 Average mutual information (AMI) values for the first 20 time lags of the synthetic IKI data. A first local minimum is visible at lag 3

series is the result of chopping off several data points equal to $t*(m-1)$ data points from the original time series, selecting the first local minimum saves data points for the actual analysis. For our example data, the first local minimum occurs at time lag 3, hence we can specify $t=3$ as the delay parameter.

After having specified the delay parameter t, we can now determine the embedding dimension parameter m using the false nearest neighbors' function (Kennel et al., 1992). The function provides an estimate of how distances between neighboring coordinates in phase-space change as a function of embedding dimension. Once these changes are small, we assume to have found a stable, appropriate dimensionality for our data.

Determining the embedding dimension parameter m is in many ways like determining the delay parameter t: we will create a plot showing the amount of false nearest neighbors for multiple embedding dimensions and use the shape of this function to help us decide which embedding parameter to choose. Figure 6 shows the false nearest neighbors' function for the first 20 embedding dimensions of our synthetic IKI data. To select the embedding dimension parameter m, we are interested in finding the first value after which the function does not change appreciably anymore. In our example, the false nearest neighbors' function drops to 0 at $m=5$ and remains stable afterward. However, sometimes the number of false nearest neighbors increases again. In that case, it is advisable to choose the first local minimum, similar to the average mutual information function (Wallot, 2017). Again, further increase of the embedding dimension comes at the cost of losing data points for subsequent analysis. Should the shape of the false nearest neighbors' function (or the average mutual information function) look particularly different from what to expect, i.e., without a clear first local minimum or flattening after an initial drop, it is always recommended to check the underlying data.

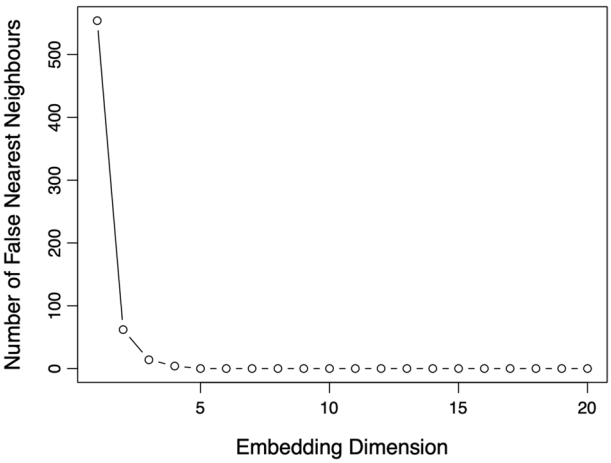

Fig. 6 False nearest neighbor values for the first 20 embedding dimensions of the synthetic IKI data. The function levels off at dimension 5

For a practical discussion of parameter selection, see Wallot (2017) or Wallot and Grabowski (2019).

Finally, we need to select a threshold parameter *r*. Recall that the threshold parameter *r* provides the width of the tolerance range within similar data points—or phase-space coordinates—which are counted as recurrent. Looking back at Eq. (2), the recurrence rate increases as we increase the radius. The choice of a threshold parameter is usually arbitrary since the results of recurrence analysis are relatively scale-invariant and robust across different radii. However, recurrence analyses are in general most sensitive when the threshold parameter *r* is set to a low, but not too-low value, so that we end up having enough recurrence points to map out the dynamics of a time series while not revealing too many recurrence points so that these dynamics 'drown' in meaningless recurrences between vastly distant data points. Figure 7 shows different recurrence plots of the synthetic IKI data for different values of the threshold parameter *r*, illustrating that with increasing radii more and more data points are counted as recurrent. As shown in Fig. 7a, hardly any recurrence structure can be identified if the radius is set too low. On the contrary, if *r* is chosen too high, we struggle to disentangle meaningful from meaningless recurrence patterns (Fig. 7c).

When selecting a radius value, it is recommended that the resulting recurrence rate lies between 1 and 5%. This range can be lower for long time series with little noise and strong deterministic components, but also higher—between 5 and 20%—for very noisy or noise-type time series (Wallot, 2017; cf. Fig. 7b). Again, the advantage is that results obtained from recurrence-based analyses are relatively robust across a range of parameter values, including the threshold parameter *r*. When in doubt as to the robustness of the results, a parameter exploration can be performed (Wallot & Leonardi, 2018). Here, results are computed for a range of plausible parameter values and then subjected to inferential statistics. While the different parameters settings will naturally change the absolute values of the RQA measures to some extent, the magnitude and direction of the effects should remain the same.

It is noteworthy that certain circumstances allow the described parameters to be set without using estimation procedures, for instance when dealing with nominal sequences such as the 'Humpty Dumpty' text. Here, we can set the delay parameter

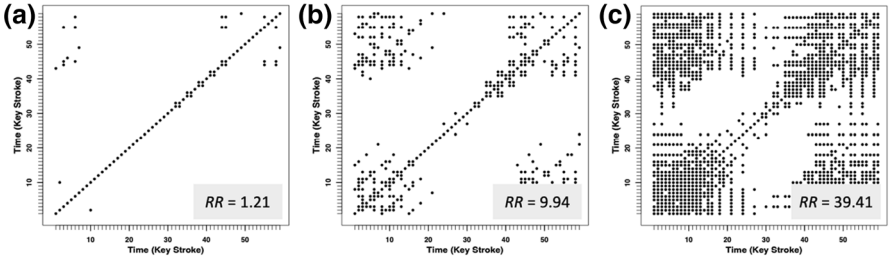

Fig. 7 Effects of the threshold parameter *r* on recurrence plots for the putative inter-key interval data and the resulting recurrence rate (RR). RQA was computed with a delay parameter $t=3$, an embedding dimension parameter $m=5$, and the threshold parameters **a** $r=0.01$, **b** $r=1.5$, and **c** $r=2.5$

as well as the embedding parameters to 1, and the threshold parameter (close) to 0. The sequence can then be analyzed in terms of its original elements, and only identical values, i.e., values belonging to the same category are counted as recurrent. However, even then there are practical reasons to choose different parameter values, for example, a higher embedding dimension to ignore recurrence patterns due to mere single letter repetitions. Moreover, additional parameters can be set to change the estimation of RQA measures (for a description see Coco & Dale, 2014). Furthermore, it is possible to normalize data before subjecting them to RQA. Data can be normalized before embedding (e.g., by z-transformation or re-scaling to unit interval [0, 1]), after embedding using the norm-parameter implemented in the RQA function (Euclidean norm, Min-norm, Max-norm, etc.), or both. Usually, it is trivial which normalization is applied to the time series or which norm is chosen for phase-space normalization as long as the same norm is applied across all compared time series.

Analysis of sample data

The steps of parameter estimation discussed so far mainly dealt with the analysis of a single time series. However, we are usually not interested in describing just a single time series, but rather in drawing statistical inferences based on the comparison of samples (of time series). This puts some extra demands on the parameter estimation procedure. While the individual steps described above essentially remain the same when dealing with sample data, we additionally must decide which parameters to use for a whole sample—or even multiple samples that should be compared.

The most basic approach is to estimate the parameters as described above for each time series within the sample (or each time series of all the samples that should be compared). Subsequently, the values obtained for the time delay and embedding dimensions are averaged and then rounded to the nearest integer value (Wallot, 2017). Thus, all time series are embedded the same way and can be compared regarding the same parameters.

To set the threshold parameter, one approach is to settle for one radius that is then applied to all time series so that the average recurrence rate across the sample equals some percentage of recurrence points, say 5%. This procedure has several advantages: On the one hand, the time series can now be compared in terms of their recurrence measures with all parameters being equal. On the other hand, the compared time series can be analyzed regarding differences in the recurrence rates as the value of the radius parameter is fixed for all time series. However, it is important to check the distribution of the recurrence rates across the sub-samples. As described above, if the recurrence rate strives towards 0%, other recurrence measures cannot be computed. If the recurrence rate aims towards 100%, recurrence measures become meaningless. While it can be tolerable if very few time series have a recurrence rate close to 0 or 100%, the majority needs to have recurrence rates such that a proper quantification of the time series is possible. If that is not the case, several options have been proposed. One option is to define a minimum threshold, e.g., 1%, for time series with recurrence rates close to 0%, so that all time series below the

specified threshold are recalculated with a radius that brings them just to this minimal percentage of recurrence points.

An alternative approach to set the threshold parameter is to fix recurrence rates to a specific value, e.g., to 5%, and adjust the radius for every single time series of the sample accordingly (Wallot et al., 2012). A drawback of this approach is that recurrence rates can no longer be a variable of interest in comparing the (sub-) samples since it is kept constant.

Further readings

For a thorough theoretical and formal treatment of recurrence-based analysis, we recommend the overview article by Marwan et al. (2007). Further, Webber and Zbilut (2005) provide a conceptual introduction to the analysis. Tutorials for how to run RQA in R (Coco & Dale, 2014; Wallot, 2017; Wallot & Leonardi, 2018) and MATLAB (Wallot & Grabowski, 2019) guide through the practical parts of parameter estimation and application of RQA in more detail. The webpage "Recurrence Plots and Cross Recurrence Plots" by Marwan and colleagues (http://www.recurrence-plot.tk) hosts a huge bibliography of research that employed recurrence-based analysis, as well as it gives brief introductions and tutorials, and lists different software packages that implement RQA.

Applying RQA to keystroke logging data of first and foreign language writing

The present study aims to demonstrate that RQA is a valuable resource for analyzing keystroke logging data that yields reliable and meaningful information about the effort a writer must exert during text production. Therefore, we will apply RQA to the keystroke logging data of first and foreign language writing and relate outcomes to conventional fluency measures. Moreover, we will explore how the information RQA provides relates to general typing skills and different levels of language proficiencies. Thereby, we seek to answer the following research questions:

(1) Do the most relevant RQA measures detect language differences between first and foreign language writing? If so, RQA measures should be higher, thus most structured, when writing in a well-known L1 and lower when writing in EFL.
(2) How do more conventional fluency measures such as variables of pausing, revising, and burst productions relate to outcomes of RQA in foreign language writing contexts? Writing in a foreign language should be more effortful and thus cause less fluent writing. Accordingly, conventional fluency measures should relate to outcomes of RQA such that more fluent writing relates to higher values of RQA measures.
(3) Does language proficiency, besides typing skills, also contribute to outcomes of RQA? RQA uses IKIs to detect global fluctuation patterns within time series data. Thus, more proficient and skilled typing should also relate to more structure

in the data and higher values of RQA measures. Accordingly, if language proficiency is a constraint that delimits a writer's performance, then higher language proficiencies should relate to higher values of RQA measures.

Participants

Forty students (19 female, mean age = 23.8 years, age range = 19–34, *SD* = 3.04) at Leibniz University of Hannover participated in the study and were each paid an expense allowance of 20€ in form of a gift card. All participants have learned German before the age of four years and received their entire education in Germany. For all participants, English was considered a foreign language that was mainly learned in school with no or very little contact before school enrollment.

Writing tasks

We developed two prompted writing assignments, both in either language. To reflect language abilities via written text production, the prompts were designed to emphasize linguistic knowledge by reproducing simple and familiar everyday subjects and procedures: How to cook your favorite meal, and how your flat looks like. Thus, prior knowledge differences regarding topic familiarity should be excluded as far as possible. By describing two different features, a known/visible object and procedure, we aimed to largely reduce learning effects across the two writing tasks while retaining the comparability of each task design. Both tasks allow to minimize planning processes but promote spontaneous and fluent writing behavior. Further, each of the two task designs is commonly considered a simple task type since it uses solely existing knowledge and hence can be reliably used to reflect language proficiency (e.g., Jost, 2021).

We realized the implementation of the tasks via written language input (for the specific writing tasks see "Appendix 1"). Since all participants were students of the academic track where reading is a common requirement and participated in an English placement test, we assumed general reading abilities to not interfere with writing skills.

Design and procedure

Data collection took place in sessions of approximately 90 min with one or two participants at the same time. After an oral instruction about the organization of the experiment, participants first completed a short questionnaire to assess a sociodemographic profile, asking for their gender, age, language(s) spoken at home and learned in school, residence abroad, duration of learning English, the weekly number of hours of English classes in school, and the use of English in every-day life. The questionnaire was developed in German and administered via pen and paper.

To evaluate language proficiency levels, participants then completed a standardized cloze test, administered by the Leibniz Language Center (LLC) of Leibniz University of Hanover. The test consists of five authentic texts with a total of 20 to 25

gaps, which must be completed. For a detailed description of the test design see http://www.c-test.de/ (Raatz, 2007–2021). The average time given per text was five minutes, resulting in a maximum of 25 min to complete the test.

To compare writing processes between German and English, a within-subject design was used in which each student first worked on the German and then on the English writing assignment. The order of tasks (i.e., describing a flat and meal preparation) altered between participants. The prompts were printed on a DIN A4 sheet in 24pt size and handed to the participant. No time restrictions and no specifications of text length were made. Participants first read the respective assignment and were then handed a notebook with an open Microsoft Word interface for task completion and an integrated keyboard in QWERTZ layout. To assess general typing skills, participants performed two copy tasks that are part of the analysis of Inputlog, developed by Luuk van Waes and Marielle Leijten (University of Antwerp, Belgium); one for each language, German (developed by Esther Breuer, University of Cologne, Germany) and English (developed by Lisa Fontaine, Cardiff University, UK & Mark Torrance, Nottingham University, UK). The copy tasks for each language were assessed before the respective writing task (for a detailed description of the task see Van Waes et al., 2019).

Data collection and analysis

Keystroke logging data was recorded with Inputlog 8 (www.inputlog.net—Leijten & van Waes, 2013). Inputlog continuously and unobtrusively logs mouse and keyboard events, as well as character position, actual document length, and copy/paste/move actions in Microsoft Word with a unique time stamp (ms). The log files were individually time-filtered using the pre-processing time filter of Inputlog to remove noise at the beginning and the end of each process and analyze the actual text production from the first to the last character or revision only. The resulting data set was used for further analysis.

For a time-based comparison of the writing processes in German and English, the integrated pause analysis of Inputlog was used with a pause threshold of 2000 ms. Pause thresholds are usually set at this level to focus on higher-level cognitive activities and bursts need to be defined at a higher process level (Van Waes & Leijten, 2015). For a product- and process-based comparison between languages, Inputlog's integrated summary and revision analysis was used.

Results of the copy tasks were analyzed using the integrated copy task analysis of Inputlog. The median IKI of the tapping task (the first out of seven subtasks) was used as an indicator of low-level typing and motor skills and the median IKI of all targeted bigrams within the copy task was used as an indicator of general typing ability.

To evaluate participants' English language proficiency as assessed via the standardized cloze test, participants automatically received points for each semantically, orthographically, and grammatically correctly filled gap. The total point value was related to the maximum possible point value with cut scores for their achieved

language level after the Common European Framework of Reference for Languages (CEFRL).

We analyzed all data in R (R Core Team, 2021). RQA was performed using the *crqa*-package in R (Coco & Dale, 2014). Keystroke logging protocols were generated using the integrated general analysis of Inputlog. The pause time between each keystroke served as the target event for RQA. As described above, the delay and the embedding parameters were estimated based on the average mutual information and false nearest neighbors functions for each time series (see "Appendix 2"). Averaging the obtained values resulted in a delay parameter $t = 1$ and an embedding parameter $m = 7$. Furthermore, we chose a threshold parameter $r = 300$, yielding an average recurrence rate between 1 and 5% across time series. After, we computed the most common RQA measures: recurrence rate, determinism rate, average diagonal line length, maximum diagonal line length, entropy, laminarity, and trapping time. Data and code is available at: https://doi.org/10.6084/m9.figshare.19794673.

Results

In the following section, we apply RQA to keystroke logging of first and foreign language writing data. Thereby, we aim to demonstrate that it provides meaningful data and (1) is a powerful method to differentiate between first and foreign language writing and (2) is sensitive to individual differences of language proficiency. To create a first perspective on the logged data, we start with a more traditional approach to analyzing the data by comparing writing process parameters such as pausing behavior, burst productions, and revision behavior for both, German and English. We then extend the analysis to RQA by comparing first and foreign language writing with well-established measures of RQA. After, we describe RQA from a multi-dimensional perspective by examining first how outcomes of RQA are related to more conventional parameters of time-series data and second how RQA is affected by general typing abilities and English language proficiency in foreign language writing contexts.

Time-, product- and process-based analyses

The study followed a two-factorial design with language (German vs. English) as a factor within participants and task assignment (German: meal—English: flat vs. German: flat—English: meal) as a factor between participants. We kept the language order of the writing assignments constant (first German, then English), and the order of task assignments alternated between participants. Mixed model analyses of variance (ANOVAs) with language as a within-subject and task assignment as a between-subject factor were conducted to compare writing processes and products of first and foreign language writing.

A time-based comparison of the writing processes in German and English is presented in Table 3. Participants spent significantly more time completing the writing assignment when writing in EFL. While the active writing time did not differ

Table 3 Time-based comparison of L1 and EFL writing (with $df = 1, 38$)

	German		English		F	p	η_p^2
	M	*SD*	*M*	*SD*			
Process time (s)	402.46	166.73	477.34	261.84	5.82	**0.02**	0.03
Active writing time (s)	318.47	133.67	331.35	196.41	0.28	0.60	0.00
Total pause time (s)	83.99	50.90	145.98	93.10	33.23	**< 0.001**	0.15
Proportion of pause time	0.20	0.08	0.31	0.11	57.82	**< 0.001**	0.22

Significant results (with $p < .05$) are formatted bold

remarkably between languages, participants paused overall longer when writing in EFL which is also reflected in the proportion of pause time. Participants' time needed for task completion, their time actively spent on writing, and their overall and proportion of pause time did not differ significantly between both writing tasks, i.e., describing a flat and meal preparation.

Regarding the final texts, participants used similar numbers of keystrokes, characters, words, and sentences to complete the writing assignments in either language. Further, participants used comparable numbers of keystrokes, characters, words, and sentences for either task assignment, indicating that both tasks worked well in either language.

Table 4 displays a comparison of language-dependent differences between conventional measures used to analyze keystroke logging data such as parameters for pausing, revising, and burst production in L1 and EFL. Typically for writing in a foreign language, participants paused more often and longer when writing in English. Accordingly, participants also produced shorter bursts with fewer characters typed between two pauses. Revision behavior did not differ between languages and seem

Table 4 Comparison of conventional writing process parameters for L1 and EFL (with $df = 1, 38$)

	German		English		F	p	η_p^2
	M	*SD*	*M*	*SD*			
Pausing behavior							
Pauses per minute*	2.88	1.05	3.73	1.08	37.23	**< 0.001**	0.15
Median pause length (s)*	3.3	0.82	3.67	0.67	8.86	**< 0.01**	0.06
Burst production							
Median duration of p-bursts (s)*	13.71	7.75	8.21	3.23	31.62	**< 0.001**	0.19
Median length of p-bursts (char)*	45.24	38.87	23.56	23.35	18.58	**< 0.001**	0.14
Revision behavior							
Revisions per minute	7.45	3.77	7.48	4.25	0.00	0.96	0.00
Deletions per minute	5.92	2.54	5.85	2.63	0.05	0.83	0.00
Efficiency	0.84	0.07	0.86	0.07	3.43	0.07	0.02

Significant results (with $p < .05$) are formatted bold

*With a pause threshold > 2000 ms

to reflect language-independent writing skills rather than specific characteristics of first and foreign language writing. Participants similarly often went back to revise already produced text by inserting or deleting parts of it. As a measure of efficiency, we calculated the ratio of all characters that the students typed while completing the task and the total amount of characters in the final text. Neither the total number of deletions nor the efficiency with which a writer compiled a text differed remarkably between languages.

Beyond the general language effect, we found a group × language interaction effect for the times a writer paused per minute ($F(1,38) = 10.79$, $p < 0.01$, $\eta_p^2 = 0.05$) ($F(1,38) = 9.64$, $p < 0.01$, $\eta_p^2 = 0.05$) and the median time ($F(1,38) = 4.61$, $p < 0.05$, $\eta_p^2 = 0.03$) and length ($F(1,38) = 4.9$, $p < 0.05$, $\eta_p^2 = 0.04$) of a burst production between two pauses. When describing a meal, participants produced more pauses and shorter bursts with on average fewer characters typed between two pauses, when writing in a well-known first language. While describing an apartment in a first language prompted more fluent writing as compared to describing a meal in L1, the effect does not apply when participants wrote in EFL.

To compare general typing skills in German and English, we calculated repeated measurement ANOVAs with language as a within-subject factor. Participants' median IKIs for all targeted bigrams of the copy task were significantly higher ($F(1,39) = 135.28$, $p < 0.001$, $n_p^2 = 0.07$) when typing in English ($M = 162.13$, $SD = 31.99$) as compared to German ($M = 146.0$, $SD = 28.35$). Hence, participants typed faster when writing in a well-known first language, but slower when writing in a foreign language. Low-level typing motor skills were assessed via the same task in either language (i.e., pressing two keys with alternating hands) and hence did not differ between languages.

Recurrence quantification analysis (RQA)

We extended the analysis of keystroke logging data to RQA following the same study design. We conducted mixed model ANOVAs for the most relevant RQA parameters with language as a within-subject and writing assignment as a between-subject factor (Table 5). The comparisons of RQA variables between languages revealed a conclusive picture: All seven previously defined parameters differed significantly between languages. Values of the German texts were overall higher, indicating more structured time series patterns when writing in a first language as compared to writing in a foreign language. Reliability analysis to examine whether the parameters can be considered a scale showed internal consistency measures with Cronbach's alpha of 0.81 for the German and 0.78 for the English texts. By applying RQA to keystroke logging data, it is hence possible to effectively distinguish between writing in a first or foreign language.

Beyond the general language effect according to which L1 writing has stronger auto-correlation structures than EFL writing, two out of seven parameters showed a group effect (recurrence rate: $F(1,38) = 6.16$, $p < 0.05$, $n_p^2 = 0.12$; average diagonal line length: $F(1,38) = 5.45$, $p < 0.05$, $n_p^2 = 0.11$) and a language × group interaction effect (recurrence rate: $F(1,38) = 9.99$, $p < 0.01$, $n_p^2 = 0.04$; average

Table 5 Comparison of L1 and EFL writing for most relevant RQA parameters (with $df = 1, 38$)

	German		English		F	p	η_p^2
	M	*SD*	*M*	*SD*			
Recurrence rate	13.92	9.72	6.99	6.62	56.02	**< 0.001**	0.17
Determinism rate	90.45	5.58	84.17	8.06	51.92	**< 0.001**	0.18
Average diagonal line length	4.91	1.37	3.93	0.99	47.51	**< 0.001**	0.16
Maximum diagonal line length	35.65	19.83	22.58	10.63	43.30	**< 0.001**	0.19
Entropy	2.13	0.38	1.80	0.37	63.91	**< 0.001**	0.19
Laminarity	77.45	14.55	59.25	23.07	43.30	**< 0.001**	0.19
Trapping time	5.75	1.89	4.28	1.46	45.88	**< 0.001**	0.17

Significant results (with $p < .05$) are formatted bold

diagonal line length: $F(1,38) = 7.1$, $p < 0.05$, $n_p^2 = 0.03$). Further, a group effect, but no group × language interaction effect was detected for entropy ($F(1,38) = 5.42$, $p < 0.05$, $n_p^2 = 0.11$). Thus, time series patterns were more structured when describing an apartment as compared to describing the preparation of a meal. This is interesting as effects of this kind were not detected within comparisons of the more conventional writing process variables which showed more pronounced effects for the German writing task only.

RQA in foreign language writing contexts

In the previous sections, we have shown that writing in an L1 and writing in EFL yielded specific and meaningful differences for writing process variables such as pausing behavior and burst production. RQA reflected these differences regarding the indicated degrees of structuring in the analyzed IKI data. In a second step, we will describe RQA in foreign language writing contexts from a multidimensional perspective. Therefore, we first compared individual outcomes of RQA to more conventional variables of writing process data. After, we aggregated the information RQA provides using a principal component analysis (PCA) that resulted in one factor. We then used its factor scores to determine the effects general typing abilities and language proficiency have on RQA's variance.

How RQA relates to conventional variables of writing process data

Table 6 presents a correlation matrix for outcomes of the individual RQA parameters and more conventional variables of the English writing process data. Regarding the time-based variables, RQA parameters did not vary with the overall time a writer needed for task completion nor the overall time a writer paused and actively wrote, indicating that different times for task completion do not affect outcomes of RQA. For relative pausing behavior, burst productions, and revision behavior, measures of RQA yielded similar patterns such that more fluent writing (as indicated by the proportion of pause times, pauses per minutes, time, and length of bursts, or

Table 6 Correlation matrix of standard writing process variables and RQA parameters for the English writing task

	RR	DET	ADL	MDL	LAM	ENTR	TT
Time-based variables							
Process time	−0.01	0.04	0.01	0.26	0.10	0.06	0.04
Total pause time	−0.24	−0.25	−0.20	−0.07	−0.12	−0.18	−0.08
Active writing time	0.10	0.17	0.11	0.38*	0.19	0.16	0.09
Prop. of pause time	− 0.37*	− 0.44**	− 0.34*	− 0.46**	−0.27	− 0.38*	−0.20
Pause behavior							
Pauses per minute	− 0.44**	− 0.44**	− 0.39*	− 0.48**	− 0.40*	−0.29	−0.24
Median length of pauses	0.11	−0.03	0.12	0.05	0.03	−0.03	0.07
Burst production							
Median duration of p-bursts	0.40*	0.40**	0.37*	0.41**	0.39*	0.25	0.25
Median length of p-bursts	0.29	0.44**	0.23	0.31*	0.35*	0.32*	0.20
Revision behavior							
Revisions per minute	0.33*	0.32*	0.26	0.40*	0.30	0.26	0.23
Deletions per minute	0.28	0.27	0.17	0.41**	0.23	0.22	0.13
Efficiency	0.21	0.24	0.24	0.17	0.24	0.05	0.17
Typing							
Number of keystrokes	0.31	0.35*	0.31	0.54**	0.36*	0.34*	0.26
Median inter-key intervals	−0.62**	−0.72**	−0.59**	−0.60**	−0.62**	−0.67**	−0.55**

RR recurrence rate, *DET* determinism rate, *MDL* average diagonal line length, *ADL* maximum diagonal line length, *LAM* laminarity, *ENTR* entropy, *TT* trapping time; pause threshold > 2000 ms

*Indicates $p < 0.05$

**Indicates $p < 0.01$

the number of revisions) related to higher values of most RQA parameters. Regarding the number of keystrokes in relation to individual RQA parameters, there was a positive correlation such that higher numbers of keystrokes occured with higher values of RQA measures. However, different lengths of time series data should lead to sufficiently robust data, as for longer texts, the chances increase to find structure but also disorder. In contrast to more conventional variables such as the number of pauses or revisions, which are usually reported in relation (e.g., per minute), RQA measures are already coupled with averages or percentages across a recurrence plot. Hence, these observations might be rather third-variable effects: Individuals with higher English proficiencies produced more keystrokes and respectively longer texts (Pearson correlation: $r = 0.33$; $p < 0.05$), which is also reflected in higher values of the RQA measures for students that produced more keystrokes. Regarding IKIs, RQA measures showed strong correlations. Since RQA uses inter-key intervals to detect recurring patterns within the underlying data, shorter IKIs (thus faster typing) also corresponded to higher values of RQA.

Effects of language proficiency and typing skills on RQA

To aggregate the information RQA provides, we conducted a PCA (Field et al., 2012). The relevant RQA parameters were highly intercorrelated (all correlations $r > 0.7$; $p < 0.001$). Although principally independent, different RQA parameters usually correlate highly with one another and do not dissociate different dynamic features but are all indicators of temporal regularity (Wallot, 2017). Therefore, we checked for multicollinearity and excluded the average diagonal line length. The remaining six variables revealed a determinant greater than 0.00001, which is a heuristic to avoid extreme multicollinearity (Field et al., 2012, p. 770). The data met the Kaiser–Meyer–Oklin criteria of sampling adequacy with an overall value of 0.82 and values for each variable > 0.77. Bartlett's test of sphericity was highly significant, $X^2(15) = 385.41$, $p < 0.001$, indicating that the data were appropriate to perform the PCA. An initial PCA resulted in one factor with eigenvalues of 5.253 that is above Kaiser's criterion of 1, explaining 87.5% of the variance. Table 7 shows the individual factor loadings. We used the resulting factor scores as aggregated information of the RQA for further analyses.

We assessed English typing skills via a copy task in the respective language and used the median IKI of the tapping task as an indicator of low-level typing and motor speed and the median IKI of all targeted bigrams within the copy task as an indicator of general typing ability. While faster typing abilities (i.e., lower values of IKIs) corresponded to higher values of RQA ($r = -0.50$, $p < 0.001$; Fig. 8a), low-level typing and motor speed did not significantly correlate with the RQA-component.

The students participated in a standardized cloze test to assess their English language proficiency levels. The test returned the percentage of correctly filled gaps as a continuously varying measure. In terms of the CEFRL, the resulting range (between 0.34 and 0.91) corresponded to competence levels of A2 to C1 with $M = 0.72$ ($SD = 0.14$). Relating participants' English language proficiency to the factor scores of the RQA-component revealed a medium–high correlation ($r = 0.43$, $p < 0.01$) with higher language proficiency measures corresponding to higher structuring in the data and thus higher values of RQA (Fig. 8b).

To investigate the predictive effects of typing skills and language proficiency on RQA's variance, we conducted a hierarchical regression analysis with RQA's aggregated factor scores as a response variable (Table 8). In a first step, we added two predictors for typing skills: low-level typing and motor speed (here typing speed) and general typing ability as median IKIs from the copy task. We introduced both variables as standardized scores to produce comparable regression coefficients for later analysis. The model overall explained 30.07% of the variance in the RQA component ($F(2,37) = 7.95$, $p < 0.01$) with a significant predictive effect of typing

Table 7 Factor loadings of the principal component analysis

	RR	DET	MAXL	ENTR	LAM	TT
Factor loadings	0.951	0.940	0.897	0.990	0.914	0.919

RR recurrence rate, *DET* determinism rate, *MDL* average diagonal line length, *LAM* laminarity, *ENTR* entropy, *TT* trapping time

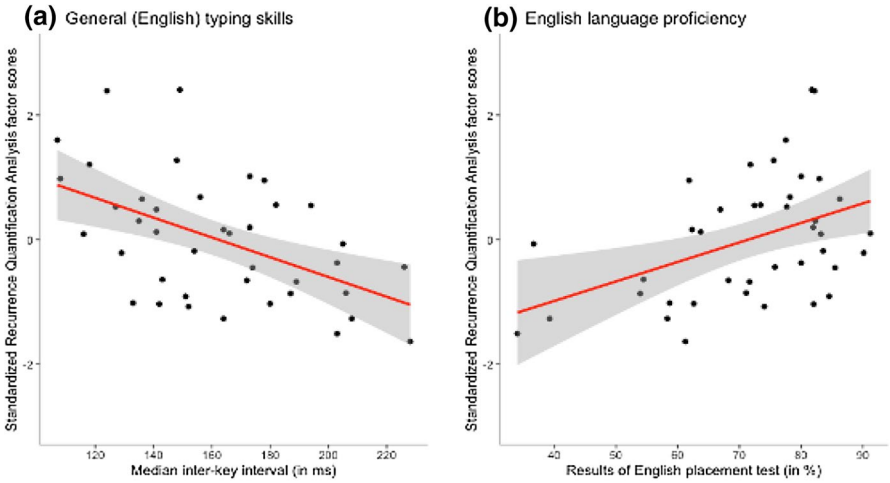

Fig. 8 Regression lines with confidence intervals for the aggregated information of the recurrence quantification analysis applied to the keystroke logging data of the English writing task show a relation to **a** the median inter-key intervals of all targeted bigrams within the English copy task used as an indicator for general typing skills and **b** the results of the English placement test as an indicator for English language proficiency

Table 8 Regression results using the RQA component as the response variable

	b (SE)	95% CI	*t*	*p*
Step 1				
Intercept	0.00 (0.14)	[−0.28, 0.28]	0.0	1
Typing ability	−0.62 (0.16)	[−0.93, −0.30]	−3.97	**< 0.001**
Typing speed	0.23 (0.16)	[−0.08, 0.55]	1.5	0.14
Step 2				
Intercept	0.00 (0.13)	[−0.26, 0.26]	0.0	1
Typing ability	−0.51 (0.16)	[−0.82, −0.19]	−3.25	**< 0.01**
Typing speed	0.28 (0.15)	[−0.02, 0.58]	1.87	0.07
Language proficiency	0.32 (0.15)	[0.02, 0.61]	2.18	**< 0.05**

Significant results (with $p < .05$) are formatted bold

CI = confidence interval

ability. Adding language proficiency measures as standardized values to the model significantly increased the explained variance of the RQA component to 38.19% ($F(3,36) = 7.41$, $p < 0.001$) with significant predictive effects of language proficiency and general typing ability. However, typing speed had no predictive effect. Therefore, we ran a model comparison between a simpler model with only typing ability and language proficiency as predictor variables and the complex model including typing speed as a predictor variable. Comparing the Akaike Information Criterion for both models ($AIC_{complex} = 103.25$; $AIC_{simple} = 104.98$) and performing

a likelihood ratio test ($\chi^2(1) = 3.72$, $p < 0.05$) both confirmed the more complex model including typing speed as a predictor variable to be a better fit. According to the model, both, typing skills and language proficiency thus explained the aggregated information of RQA's variance in foreign language writing contexts.

Discussion

In this paper, we introduced recurrence quantification analysis (RQA) as a global dynamic analysis for time series of inter-key intervals (IKI) that tracks recurring patterns within writing processes. We have shown that RQA produces data that reliably reflects differences between first and foreign language writing, thus scrutinizing its potential when applied to research questions of foreign language writing. By inspecting RQA from a multi-dimensional perspective, we further demonstrated that (1) RQA produces robust data across different lengths of time series, (2) higher values of RQA tend to correspond to more fluent writing, and (3) both, typing skills and language proficiency contribute significantly to RQA's variance, thus comprising different sources of the observed IKI progression during writing.

Foreign language writing and RQA

Foreign language production is effortful and requires extra attention during the process of writing. The present data demonstrate that variables of both methodological approaches reflect these differences in effort: Participants paused significantly more and longer, and produced overall shorter bursts with fewer characters when writing in EFL. Accordingly, we also observed more effortful language production in outcomes of RQA measures when writing in a foreign language: As expected, time series patterns of text production were distinctively more structured when writing in L1 as compared to EFL. Further, variables of pause and burst productions of the English writing data indicating more fluent writing also correspond to higher values of individual RQA measures.

Conceptually, this fits the hypothesis that skill constrains writing performance: Language and motor skills should act to functionally constrain the typing dynamics such that typing is more efficient for skilled writing. Since all RQA measures except recurrence entropy reflect aspects of dynamic structure, we expected all of these to be higher for skilled typing. For well-sampled continuous data, the different measures can potentially be differentiated and capture unique aspects of the underlying dynamics of the data (Marwan et al., 2007). However, for highly stochastic data or inter-event-time series such as IKIs, these measures correlate very highly—in our sample, the intercorrelations are between $r = 0.73$ and 0.98. Thus, these measures tend to tap into the same, overarching degree of temporal structure of the time series.

An exception is recurrence entropy. Intuitively, entropy seems to indicate ambiguity, and should hence be negatively correlated with measures of temporal structure. However, we are not determining the entropy of the raw signal, but the entropy of the clusters of recurrence. Thus, recurrence entropy rather reflects complexity than

ambiguity or lack of structure. Complexity implies variation of structure and can in the present context be interpreted as the adaptive flexibility with which skilled writers type. This would complement the skill-equals-constraints hypothesis in the sense that higher skill should reduce only unwanted variability but increase flexible adaptivity of behavior at the same time (Riley & Turvey, 2002). However, this is only a conjecture that needs further investigation.

Comparing revision behavior and the efficiency with which writers composed a text in L1 and EFL, however, did not reveal language-specific differences, but rather seem to reflect language-overarching writing strategies. Formulating complex, grammatical sentences and using appropriate vocabulary is cognitively more demanding and requires extra attention and surveillance when writing in a foreign language and for writers with lower language proficiencies (Lindgren et al., 2008; Van Waes & Leijten, 2015). Considering participants' overall relatively high language proficiency it is likely that language-specific differences regarding revision behavior of linguistic form did not emerge.

The study used a within-subject design in which each participant wrote two texts in either language, first in German and then in English. To largely reduce learning effects while maintaining the comparability of each task design, we developed two prompted writing assignments that featured the description of two different aspects: a known procedure, the preparation of a meal, and a known and visible object, the flat the participants currently lived in. Although both tasks neither required prior knowledge regarding topic familiarity nor complex knowledge regarding text structure, both analysis approaches revealed task-specific differences in which describing an apartment tended to prompt more fluent writing in a first language than describing the preparation of a meal. Describing the preparation of a meal usually follows specific steps that need to be carefully remembered, reflected in more interruptions of text production. In contrast, describing an apartment can be approached from various directions and thus might have been more easily accessible for the participants (see already Linde & Labov, 1975). While the conventional variables of keystroke logging analysis such as pausing behavior and burst production did reflect these systematic differences only for L1 writing, recurrence rate, the average diagonal line length, and the recurrence entropy do show, although more pronounced in L1, significant differences between the two task designs for either language. Consequently, RQA detects systematic differences between the two different writing assignments and their cognitive and linguistic demands which the more conventional measures only reflect when writing in a well-mastered first language. This is in line with considerations regarding skill, constraint, and flexible adaptivity of behavior.

Relating individual measures of RQA to time-based and more conventional writing process variables scrutinizes one of its strengths. Measures of RQA did not covary with time-based variables such as the processing time, the overall pausing time, the active writing time, or the median length of a pause, indicating that different lengths of time series data also led to sufficiently robust data. The longer a writer composes a text the higher are the chances to find structure in the data. Similarly, the chances increase to find disorder. Recall that, in contrast to more conventional measures, RQA measures are already coupled with averages or percentages across a recurrence plot. So, if the observation period is sufficiently captured then RQA also

captures the relevant processes comparably in data of different time lengths. Conventional variables strongly vary with different text lengths (e.g., the longer a writer composes a text, the higher are the chances of long pauses or more revisions) which makes it difficult to compare texts of different lengths. In contrast, RQA produces robust data across different lengths of time series as long as the relevant processes are sufficiently captured (e.g., not comparing a text of only a few words to texts of 200–300 words). In line with our expectations to find more fluent and skilled writing revealing also more structured typing time series patterns, relative measures such as number of pauses, bursts, or revisions per minute, and median time and length of a burst between pauses do vary with (most) individual measures of RQA such that more fluent writing goes along with more detected recurring patterns, more structured data, and hence higher values of RQA measures.

RQA uses IKIs to detect recurring patterns within the writing data. Hence, it is not surprising that, on the one hand, shorter IKIs and respectively faster typing correspond to higher values of RQA, and, on the other hand, general typing skills are a reliable predictor of RQA's variance. According to our regression model, however, RQA also reveals information about language skills such that participants with higher language proficiencies produced more structured data resulting in higher values of RQA. Following from that, the present data support the idea that RQA yields useful information about the constraints (e.g., language constraints) a writer has during text compilation.

The focus of interest of more conventional approaches to researching time series of IKIs are primarily those writing phases in which the graphomotor activity is interrupted, thus paused during writing (Weinzierl & Wrobel, 2017). Accordingly, constraints of the cognitive-motor system, as introduced above, might be complemented by a component-oriented view: Typing series usually consist of many intervals of short durations that mainly reflect the motor execution of finger movements. Thus, longer intervals (i.e., interruptions) could reflect problems related to linguistic competency. In terms of recurrence analysis, this could also lead to less regularity in foreign language typing. However, it is rather difficult to specify the processes involved in producing a single, prolonged key-press interval. Variability in typing time series could also be introduced by lack of typing skill (e.g., two-finger typing vs. ten-finger typing) or attention and not only by problems of language processing. Such an investigation would require measurements faster than a key-press in order to resolve the processes that happen between two consecutive key presses. Literature investigating fluctuations and autocorrelation structures in tapping and simple response times suggests that it is probably not warranted to conceptualize simple motor responses as locally determined (for a discussion see e.g., Ihlen & Vereijken, 2010; Kuznetsov & Wallot, 2011; Van Orden et al., 2003). While an interpretation in terms of the contribution of specific cognitive components causing regularity in typing times expressed in terms of recurrence measures is entirely possible, we have the intuition that the merit of their application to typing data lies in the abstraction from the multitude of concrete processes that might compose variability of typing data in writing tasks. RQA provides a means to capture variability and structure in time series that might have risen from any of these sources. The notion of

motor-cognitive constraints on performance provides a first bridge to link such patterns to skill and competency in typing performance.

Prospects and limits of RQA

RQA is a model-free technique such that no model or function is fitted to the data. The downside of this is that, in order to draw inferences from RQA results, either bootstrapping procedures need to be applied (Schinkel et al., 2009) or outcome variables must be submitted to inferential statistics as shown in this paper. Moreover, the selection of parameters and the use of the multiple outcome measures can be cumbersome and might sometimes warrant exploration of the parameter space (i.e., trying out multiple plausible combinations of embedding parameters to check for the stability of the results), or some form of summarizing the multiple outcome measures (see Wallot & Leonardi, 2018, for a summary of the current best practice).

However, since RQA is model-free, it is well suited for data with heterogeneous temporal variance—such as IKIs—and can be applied without having to define specific criteria of features within the data, like minimum pause times, that need to be extracted. Many of the currently used process measures rely on arbitrary thresholds (like pause thresholds to define pauses; cf. Van Waes et al., 2021) and usually require some linguistic analysis (i.e., which typing time belongs to a word, which time belongs to a sentence). While, admittedly, the threshold parameter of RQA is also chosen somewhat arbitrarily, it still yields similar results across a large range of values.

Regarding the embedding parameters, particularly the embedding dimension, the discrete inter-event-series analyzed here (series of IKIs) are point-wise samples from a continuous process. This is a cyclic kind of movement of the fingers pressing down on keys of a keyboard. The IKIs, however, conflate timing of movements with other aspects of this dynamic process, e.g., which finger had been used to execute the movement, and how the hand needed to move horizontally to allow for such a key press. The embedding dimension estimated for the present data might thus not equate the dimensionality of the more continuous movements of the fingers. If one were able to measure such movements, multivariate extensions of RQA (Wallot et al., 2016) could be used to quantify the multidimensional finger movement profiles of typing as well, and compare them to the results of IKI data.

RQA can add global information about the temporal dimension, ideally providing information complementary to the established measures of the typing process. Our results show that RQA measures seem to be more sensitive towards interactions between language and task compared to traditional measures. Moreover, they yield a non-linear relationship between foreign language writing skill and recurrence properties, potentially reflecting even more sources of effects, like the individual typing style, that traditional measures do not pick up.

As RQA can be readily applied to typing time series to extract dynamic properties, it might further prove to be a means to connect different writing processes (e.g., keystroke intervals and hand-writing trajectories). Moreover, RQA measures are connected to other linguistic-cognitive process measures like response times or eye

movements during reading. This might allow for new, comparative ways to investigate these different processes. However, this warrants a broader application of RQA in the field of writing research to better understand how to interpret RQA measures in terms of writing dynamics and to gain a more profound knowledge of how they could complement the measures currently used to describe the writing process.

Here, we mainly introduced RQA as an analysis tool that taps into the global dynamics over a whole typing task. This is in line with our considerations regarding skill and constraints on writing performance since constraints cannot be found in individual data points but must be quantified across a coherent set of data points. However, RQA is a very versatile technique, and can in principle be used to investigate local fluctuations of the writing process such as within-word dynamics using a variation called windowed RQA (Marwan et al., 2007; Wallot, 2017).

Further research

So far, RQA measures are not integrated into current linguistic, psychological, and motor-control theories of writing. To integrate these measures in a meaningful way, more studies applying RQA to the writing process are needed to gain a broader empirical understanding of what these measures contribute and how they vary. Moreover, RQA is not yet tightly integrated into our conceptual understanding of how writing works. However, originating from dynamic systems analysis, RQA brings new concepts to the table, such as the notion of constraints and flexible adaptability, and it comes with measures to appropriately quantify writing data.

Previously, we have hinted at the fact that skill is a multifaceted concept. Thinking in terms of dynamical systems, we can now propose that skill controls writing in the way that it functionally constrains behavior, but below that level, its different facets must be elaborated. This could be tested by tracing back the quality of writing performance to individual performance measures (e.g., speed, burstiness, pause patterns) which are shaped by constraints that in turn can be captured by measures of RQA and then linked to global measures of aptitude. Admittedly, at this point, this is just a speculative research agenda. However, applying RQA to implement this, and using, for instance, a path-model like approach could provide an integrative view across the different levels of control that drive the writing process.

Appendix 1: Writing tasks in German and English

Language	Writing task
German	Was ist dein Lieblingsessen, das du selbst kochen kannst? Beschreibe, wie du dabei vorgehst
	Beschreibe die Wohnung, in der du gerade wohnst
English	What is your favorite dish you can prepare yourself? Describe how you make it
	Describe the flat you live in

Appendix 2: Average mutual information and false nearest neighbors functions

See Fig. 9.

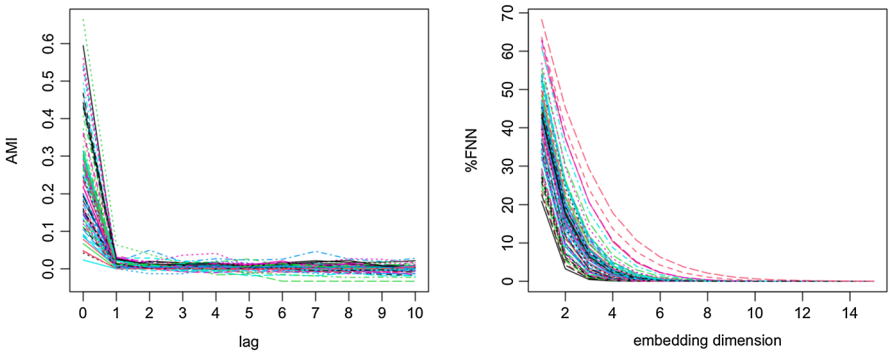

Fig. 9 The average mutual information (AMI) function (left) and false nearest neighbors (FNN) function (right) for each time series revealed a delay parameter $t = 1$ and an embedding parameter $m = 7$

Acknowledgements The manuscript has not been published previously and is not under concurrent consideration. All authors listed in the byline contributed to the study and have agreed to the order and to submission of the manuscript in its current form. The corresponding author takes the responsibility for keeping the co-authors informed of our progress through the editorial review process, the content of the reviews, and any revisions made.

Funding Open Access funding enabled and organized by Projekt DEAL. The research was funded by the affiliations of the authors.

Data availability Data and code are publically available at: https://doi.org/10.6084/m9.figshare.19794673.

Declarations

Ethics approval My co-authors and I do not have any interests that might be interpreted as influencing the research, and applicable ethical standards, specifically APA standards, were followed in the conduct of the study.

Human and animal rights The rights of the human subjects participating in our research were protected.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80–98. https://doi.org/10.1177/0741088301018001004

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication, 20*, 99–118. https://doi.org/10.1177/0741088303253572

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology, 5*, 510. https://doi.org/10.3389/fpsyg.2014.00510

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning, 39*, 81–141. https://doi.org/10.1111/j.1467-1770.1989.tb00592.x

Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters, 4*(9), 973–977. https://doi.org/10.1209/0295-5075/4/9/004

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications Ltd.

Fraser, A. M., & Swiney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A, 33*(2), 1134–1140.

Galbraith, D., & Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing. In E. Lindgren & K. Sullivan (Eds.), *Observing writing* (pp. 306–325). Brill. https://doi.org/10.1163/9789004392526015

Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication, 23*, 135–141. https://doi.org/10.1177/0741088306286283

Ihlen, E. A., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/*f*α fluctuation. *Journal of Experimental Psychology: General, 139*(3), 436–463. https://doi.org/10.1037/a0019098

Jokinen, J. P. P., Sarcar, S., Oulasvirta, A., Silpasuwanchai, C., Wang, Z., & Ren, X. (2017). Modelling learning of new keyboard layouts. In *CHI ,17: Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 4203–4215). https://doi.org/10.1145/3025453.3025580

Jost, J. (2021). Schreibaufgaben zur Indikation von Schreibkompetenz. In M. Becker-Mrotzek & J. Grabowski (Eds.), *Schreibkompetenz in der Sekundarstufe: Theorie, Diagnose und Förderung*. Waxmann.

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT Press.

Kennel, M. B., Brown, R., & Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physics Review A, 45*, 3403.

Kuznetsov, N., & Wallot, S. (2011). Effects of accuracy feedback on fractal characteristics of time estimation. *Frontiers in Integrative Neuroscience, 5*, 62. https://doi.org/10.3389/fnint.2011.00062

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication, 30*, 358–392. https://doi.org/10.1177/0741088313491692

Lekscha, J., & Donner, R. V. (2018). Phase space reconstruction for non-uniformly sampled noisy time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 28*, 085702. https://doi.org/10.1063/1.5023860

Linde, C., & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language, 51*, 924–939.

Lindgren, E., Sullivan, K., & Spelman Miller, K. (2008). Development of fluency and revision in L1 and L2 writing in Swedish high school years 8 and 9. *ITL—International Journal of Applied Linguistics, 156*, 133–151. https://doi.org/10.2143/ITL.156.0.2034428

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*, 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

O'Brien, B. A., & Wallot, S. (2016). Silent reading fluency and comprehension in bilingual children. *Frontiers in Psychology, 7*, 1265.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raatz, U. (2007–2021). *C-Test: Der Sprachtest*. Retrieved from http://www.c-test.de/

Riehl, C. M. (2014). *Mehrsprachigkeit: Eine Einführung*. Wissenschaftliche Buchgesellschaft.

Riley, M. A., Bonnette, S., Kuznetsov, N., Wallot, S., & Gao, J. (2012). A tutorial introduction to adaptive fractal analysis. *Frontiers in Physiology, 3*, 371. https://doi.org/10.3389/fphys.2012.00371

Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of Motor Behavior, 34*(2), 99–125. https://doi.org/10.1080/00222890209601934

Schinkel, S., Marwan, N., Dimigen, O., & Kurths, J. (2009). Confidence bounds of recurrence-based complexity measures. *Physics Letters A, 373*(26), 2245–2250. https://doi.org/10.1016/j.physleta.2009.04.045

Smethurst, C. J., & Carson, R. G. (2001). The acquisition of movement skills: Practice enhances the dynamic stability of bimanual coordination. *Human Movement Science, 20*(4–5), 499–529. https://doi.org/10.1016/S0167-9457(01)00065-3

Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics, 898*, 366–381.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*(3), 331–350. https://doi.org/10.1037/0096-3445.132.3.331

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition, 38*, 79–95. https://doi.org/10.1016/j.compcom.2015.09.012

Van Waes, L., Leijten, M., Pauwaert, T., & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with Inputlog. *Journal of Open Research Software, 7*, 30. https://doi.org/10.5334/jors.234

Van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2021). Measuring and assessing typing skills in writing research. *Journal of Writing Research, 13*, 107–153. https://doi.org/10.17239/jowr-2021.13.01.04

van Weerdenburg, M., Tesselhof, M., & van der Meijden, H. (2019). Touch-typing for better spelling and narrative-writing skills on the computer. *Journal of Computer Assisted Learning, 35*, 143–152. https://doi.org/10.1111/jcal.12323

Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes, 54*, 382–405. https://doi.org/10.1080/0163853X.2017.1297921

Wallot, S., & Grabowski, J. (2019). A tutorial introduction to recurrence quantification analysis (RQA) for keystroke logging data. In E. Lindgren & K. Sullivan (Eds.), *Observing writing* (pp. 163–189). Brill. https://doi.org/10.1163/9789004392526_009

Wallot, S., & Leonardi, G. (2018). Analyzing multivariate dynamics using cross-recurrence quantification analysis (CRQA), diagonal-cross-recurrence profiles (DCRP), and multidimensional recurrence quantification analysis (MdRQA)—A tutorial in R. *Frontiers in Psychology, 9*, 2232. https://doi.org/10.3389/fpsyg.2018.02232

Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1745–1765. https://doi.org/10.1037/xlm0000030

Wallot, S., O'Brien, B. A., & Van Orden, G. (2012). Fractal and recurrence analysis of psycholinguistic data. In C. Westbury, G. Jarema, & G. Libben (Eds.), *Methodological and analytic frontiers in lexical research* (pp. 395–430). John Benjamins.

Wallot, S., Roepstorff, A., & Mønster, D. (2016). Multidimensional recurrence quantification analysis (MdRQA) for the analysis of multidimensional time-series: A software implementation in MATLAB and its application to group-level data in joint action. *Frontiers in Psychology, 7*, 1835. https://doi.org/10.3389/fpsyg.2016.01835

Webber, C. L., Jr., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary non-linear methods for the behavioral sciences* (pp. 26–94). Retrieved from https://www.nsf.gov/pubs/2005/nsf05057/nmbs/nmbs.pdf

Webber, C. L., Jr., & Zbilut, K. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology, 76*, 965–973. https://doi.org/10.1152/jappl.1994.76.2.965

Weinzierl, C., & Wrobel, A. (2017). Schreibprozesse untersuchen. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 221–237). Waxmann.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Lisa Haake[1] · Sebastian Wallot[2,3] · Monika Tschense[2,3] · Joachim Grabowski[1]

[1] Department of Psychology, Leibniz Universität Hannover, Schloßwender Str. 1, 30159 Hannover, Germany

[2] Department of Psychology, Leuphana Universität Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany

[3] Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany