Keystroke Dynamics Predict Essay Quality

Aaron D. Likens (Aaron.Likens@asu.edu)
Laura K. Allen (LauraKAllen@asu.edu)
Danielle S. McNamara (danielle.mcnamara@asu.edu)

Arizona State University, PO BOX 872111
Tempe, AZ 85281 USA

Abstract

Language entails many nested time scales, ranging from the relatively slow scale of cultural evolution to the rapid scale of individual cognition. The nested, multiscale nature of language implies that even simple acts of text production, such as typing a sentence, entail complex interactions involving multiple concurrent processes. As such, text production may have much in common with other cognitive phenomena thought to emerge from multiplicative interactions across temporal scales, namely those that exhibit fractal properties. We investigated the relationship between fractal scaling and the quality of produced text. Participants (*N*=131) wrote essays while their keystrokes were recorded. Fractal analyses were then performed on time series of interkeystroke intervals (IKIs). Results showed that fractal properties characterizing IKIs positively predicted expert ratings of essay quality, even after accounting for essay length. The results support our hypotheses concerning multiscale coordination and text production.

Keywords: text production; writing; keystroke; multifractal; essay quality

Introduction

Recent theoretical and empirical work characterizes language as a complex, dynamic system that involves the coordination of multiple nested time scales (Dale, Kello, & Schoenemann, 2016; Rączaszek-Leonardi, 2010; Rączaszek-Leonardi & Kelso, 2008). Consider three time scales that have been highlighted extensively in the literature. Language can evolve on a relatively slow scale along with changes in cultures and significant historical events. On a faster scale, language can be altered throughout an individual's life, based on their experiences and knowledge. Lastly, language can change at more rapid scales in response to cognitive events that can span days, hours, or mere fractions of a second.

These time scales span several orders of magnitude and paint a complex picture of language. The picture is further complicated because each of the time scales implicates different systems (e.g., cultural, interpersonal, physiological) and suggests that language processes should be studied as complex, dynamical systems. The current work explores this idea in the context of text-based language production. We examine whether dynamic analyses of typing behaviors during essay writing provide empirical support for the notion of writing as a complex, dynamical system.

The nested, multiscale character of text production is apparent in the simple example of typing an essay. The relatively fast time scale of word selection is nested within and constrained by the slower time scale of idea generation. Singular ideas are further nested within the subtopics and global topic of the essay that change at even slower rates. Beyond these examples, nesting can continue at both faster and slower time scales. Rapidly changing physiological processes influence and support the act of writing that would not be possible without a lifetime of learning or the evolution of a language within a culture. Thus, even the seemingly simple act of typing one sentence of an essay may entail complex interactions of any number of processes, each with its own characteristic rate of evolution. The implication is that language production involves the coordination of numerous systems over many different time scales. Our assumption is that the act of text production (i.e., typing an essay) provides a window into ongoing cognitive processes (Pinet, Ziegler, & Alario, 2016). As such, we expect that keystroke dynamics will reveal the multiply-nested character of text production.

Multiscale Interactions in Human Behavior

A wide range of cognitive phenomena have been described as emerging from the interaction of multiply-nested time scales (Ihlen & Vereijken, 2010). The principle evidence for that claim is the observation of *fractal scaling*. Fractal scaling typically refers to two qualities: long-range autocorrelation and scale dependence. Long-range autocorrelation implies that time series observations exhibit significant correlations over large timespans (Beran, 1994). That is, an observation made at one point in time is related to subsequent observations that extend into the future. Scale dependence suggests that measurements of time series (e.g., variance) depend on the temporal scale at which they are measured (Mandelbrot & Van Ness, 1968).

There are numerous examples of behavioral time series known to exhibit fractal scaling: reaction times (Gilden, Thornton, & Mallon, 1995; Van Orden et al., 2003), time estimation (Wagenmakers et al., 2004), eye movements (Stephen & Anastas, 2011), hand movements (Anastas, Stephen, & Dixon, 2011; Stephen, Arzamarksi, & Michaels, 2010), arm movements (Chen, Ding, & Kelso, 1997), postural corrections (Collins & DeLuca, 1993), and various forms of tool-use (Likens, Fine, Amazeen, & Amazeen, 2015; Nonaka & Bril, 2014).

Much of the work on fractal scaling in cognition has emphasized interaction across scales as its primary theoretical contribution (Ihlen & Vereijken, 2010; Kelty-

Stephen & Wallot, in press). The basic idea is that the rich structure observed in behavioral time series is the product of many simultaneously occurring processes (e.g., physical, cognitive). Each process exists on its own time scale, with effects of slower time scales multiplicatively cascading to faster and faster time scales. As such, fractal scaling reflects on-the-fly cognitive organization during tasks (Van Orden et al., 2003; Wallot, Hollis, & van Rooij, 2013). Further, variability in fractal properties reflects the flexibility and adaptability in typical cognitive tasks necessary for coordination across those levels (Anastas et al., 2011).

If fractal scaling reflects the flexibility and adaptability that stems from multiscale coordination, then reliable relationships should exist between fractal scaling and other meaningful aspects of behavior. The literature contains several such examples. Visual search is faster when eye movements exhibit fractal properties (Stephen & Anastas, 2011). Fractal variability in hand movements predicts better perceptual estimates (Stephen et al., 2010). Moreover, fractal patterns distinguish between various forms of skilled and non-skilled behavior (e.g., Nonaka & Bril, 2014). These examples are not exhaustive but hint at the large number of skillful behaviors that exhibit fractal characteristics.

The current work explores the idea that fractal scaling might also reflect the multiscale coordination involved in the skilled production of text. Across domains, the evidence implies being skilled means adapting to task demands, and fractal scaling characterizes flexibility (Gorman et al., 2010; Nonaka & Bril, 2014). The observation of flexibility in skilled text production (Allen, Snow, & McNamara, 2016) leads to the hypothesis that fractal scaling will reveal the flexibility required from the nested, multiscale act of composition. That is, we expect more skilled text production will be characterized by fractal variability.

We are not aware of any studies that have examined the time course of text production for evidence of fractal scaling; however, work related to reading and skilled typing provides some bases for exploration (e.g., Wallot & Grabowski, 2013; Wallot et al., 2013; Wijnants et al., 2012). For example, Wijnants and colleagues (2012) showed that the presence of fractal scaling in word naming times distinguished dyslexic and non-dyslexic readers. They also found a positive relationship between fractal scaling and reading fluency. Another study involving skilled typing suggests that fractal properties may depend on task complexity/difficulty (Wallot & Grabowski, 2013). The relevant finding in that study was that there was greater fractal variability over time when participants typed a set of directions than when they typed simple lyrics from memory or simply copied text.

Current Study

This study investigates how fractal properties in keystroke logs are related to the quality of written text. Participants wrote timed, prompt-based argumentative essays while their keystrokes were recorded. Time series were constructed from the latencies between keystrokes and analyzed by fractal analysis. Essays were scored by experts on holistic quality and analytical subscales. This study is exploratory and the first of its kind; nonetheless, our general expectation is that, like performance on other tasks, fractal properties will serve as reliable predictors of essay quality.

Method

Participants Undergraduate students (*N* = 131, Female = 58, mean age = 19.8 years) were recruited from a large university in the United States. Students participated in the study in exchange for course credit.

Procedure Participants wrote a timed (25-minutes), prompt-based, argumentative essay. Essay prompts were similar in structure to Scholastic Aptitude Test (SAT) prompts in that participants were asked to take either a supporting or contrary position on a given topic. Keystrokes and their respective time stamps were recorded while students composed their essays. Unsurprisingly, participants varied considerably in the number of keystrokes they produced (*M* = 3,385.40, *SD* = 1,107.03). To prevent bias, only the first 999 keystrokes were retained for further analysis, corresponding to lowest number of keystrokes in our sample. No other keystrokes (e.g., backspaces) were omitted. Keystroke timestamp series were then differenced to obtain time series of interkeystroke intervals (IKIs). Mouse movements were not recorded.

Text Analyses Pairs of raters evaluated the essays based on holistic quality and analytic subscales. Raters received extensive training before scoring and received compensation for their time. Holistic scores ranged from one (minimum) to six (maximum) and were based on a standardized rubric used in the assessment of SAT essays. Interrater reliability was good (*r* = 0.75). Raters were instructed to treat the distance between points (e.g., 1-2, 3-4, 4-5) as equal. The nine subscales, also based on a 6-point scales, were:

Introduction. (*M* = 3.97, *SD* = 0.96) Demonstrates mastery in meeting the goals of an introduction (e.g., presenting a topic, providing a purpose, clearly stating a thesis, previewing arguments).

Body. (*M* = 4.08, *SD* = 0.90) Demonstrates mastery in meeting the goals of body arguments (e.g., transition between arguments, using topic sentences, supporting arguments with evidence, and maintaining a flow throughout the arguments).

Conclusion. (*M* = 3.19, *SD* = 1.32) Demonstrates mastery in meeting the goals of a conclusion (e.g., summarizing the essay, re-establishing the significance of discussion, capturing the reader's attention, and effectively closing the essay).

Organization. (*M* = 3.86, *SD* = 0.98) Follows a logical structure, beginning with the introduction, through the arguments and evidence presented in the body arguments, and to the conclusion.

On-Topic/Global Cohesion. (*M* = 4.13, *SD* = 0.85) Details presented throughout the essay support the thesis and do not stray from the prompt and the main ideas and organizing principles presented in the introduction.

Grammar, Syntax, & Mechanics. (M = 3.70, SD = 0.79) Employs correct Standard American English, avoiding errors in grammar, syntax, and mechanics; the essay conveys strong control of the standard conventions of writing.

Voice. (M = 4.09, SD = 0.76) The writer is expressive, engaging, and sincere, with a strong sense of audience.

Word Choice. (M = 4.07, SD = 0.71) Word choice is precise and effective.

Sentence Structure. (M = 4.06, SD = 0.75) Sentence patterns are varied effectively, enhancing the quality of the essay.

Fractal Analysis Fractal analysis comes in two forms, monofractal analysis and multifractal analysis, both of which were performed on the IKI time series. The goal of monofractal analysis is to understand how variability depends on scale (e.g., Eke, Herman, Kocsis, & Kozak, 2002). In general, evaluating monofractality means estimation of scaling exponents from the relationship, $F^2(s) \sim s^H$ where H is the Hurst exponent, and $F^2(s)$ is a measure of fluctuation. Being a singular measure, the Hurst exponent provides a measure of typical scaling behavior in a time series. Moreover, H ranges from zero to one and has useful interpretive ranges (Collins & DeLuca, 1993; Gorman et al., 2010). When H = 0.5, the time series exhibits random variation. When H > 0.5, the series contains long-range autocorrelation, and when H < 0.5, the series exhibits long-range anticorrelation such that small values generally follow large values and vice versa. Many series have been shown to require not one but a spectrum of exponents to characterize their variability (Ihlen & Vereijken, 2010; Kantelhardt et al., 2002). Hence, the goal of multifractal analysis is to determine whether fractal scaling is fixed across time; that is, whether a time series exhibits multifractality (Kantelhardt et al., 2002).

We used Multifractal Detrended Fluctuation Analysis (MFDFA; Kantelhardt et al., 2002) to evaluate both monofractal and multifractal properties in IKIs. The outcome of MFDFA is the multifractal spectrum. MFDFA is the generalization of Detrended Fluctuation analysis (DFA) and has been used in diverse literature to characterize time-varying structure (Kantelhardt et al., 2002; Peng et al., 1994). The MFDFA procedure consists of five steps. The first step is to create the profile by integrating over a mean-centered time series. In a second step, the time series of length, N, is divided into $N_s = \text{int}(N/s)$ non-overlapping bins, such that each bin contains s observations. To compensate for N_s often being a non-integer multiple of s, the binning procedure is performed twice by starting from each end of the time series. The result partitions the time series into $2N_s$ bins. In a third step, data in each bin is fit with a least squares regression line that is subtracted from the binned data to obtain local residuals. The bin-wise residuals are squared and averaged to obtain a measure of variance within each segment, v. The fourth step averages over all the bins to obtain the qth order fluctuation function as captured in

$$F_q(s) = \{\frac{1}{2N_s}\sum_{v=1}^{2N_s}[F^2(v,s)]^{q/2}\}^{1/q}, \qquad (1)$$

where $F^2(v,s)$ is the variance calculated in Step 3 and q takes on both positive and negative values. Steps 2 through 4 are repeated for several s, increasing s by a power. The current work used a fractional power (11/10) for varying s which allowed for a larger range of scales over which scaling estimates were made. The maximum s was $\leq N/4$. Step 5 evaluates scaling behavior by performing a log-log regression of $F_q(s)$ on s for each value of q. We used 101 values of q, ranging from -3 to 3. When scaling properties are present, the result from Step 5 is a linear slope equal to the q-order Hurst exponent, $H(q)$. When q = 2, the procedure is equivalent to standard DFA. $H(q)$ can then be used to estimate the width of the multifractal spectrum $dh(q)$. In contrast H, $dh(q)$ provides a measure of the variability in scaling over time.

Results

Hierarchical multiple regression was used to explore the relations between the fractal properties in IKIs (i.e., H, $dh(q)$) and holistic essay scores (M = 3.85, SD = 0.89)[1]. Table 1 presents the descriptive statistics for the predictor variables used in constructing regression models.

Table 1. Descriptive statistics

Variable	M	SD
Number of Words (NW)	412.67	162.22
$dh(q)$	1.32	0.26
H	0.51	0.06

In addition to fractal properties, we included the total number of words (NW) in each essay as a predictor in the regression model because of the known positive relationship between essay length and essay quality (e.g., McNamara, Crossley, & Roscoe, 2013; McNamara, Crossley, Roscoe, Allen, & Dai, 2015). Predictors were checked for multicollinearity and all variance inflation factors were less than 2 (VIF_{NW} =1.09; VIF_H = 1.21; $\text{VIF}_{dh(q)}$ = 1.26), indicating that multicollinearity was not a concern. Note that, NW, H, and $dh(q)$ were converted to z-scores to aid in interpretation. This was especially crucial in the case of H as its theoretical domain is (0, 1). NW was entered in the first model step; H and $dh(q)$ were both entered in the second model step. As expected, the initial model was significant, β = 0.47, R^2 = 0.28, p < 0.001; however our interest was in characterizing whether fractal properties predicted essay quality over and above NW. The results showed that fractal properties improved model fit, $F(2,127)$ = 6.68, p < 0.01, R^2 = 0.35. As expected NW was a significant predictor such that a one standard deviation increase in essay length predicted a 0.54 increase in holistic score, $t(127)$ = 8.09, p <

[1] We also estimated models that included polynomial terms. However, none of the polynomial models improved model fit and were not reported here.

0.001. After controlling for NW, the model also revealed that a one standard deviation increase in *H* predicted a 0.21 increase in holistic score, $t(127) = 2.92$, $p < 0.01$. Furthermore, a one standard deviation increase in $dh(q)$ predicted a 0.29 increase in holistic score, $t(127) = 3.19$, $p < 0.01$.

Following the analysis of holistic essay scores, nine additional sets of regression models were fit predicting each subscale from NW, *H*, and $dh(q)$. The modeling strategy for these additional models was the same as for overall essay quality. A summary of those models appears in Table 2. The table shows that fractal properties explain significant variance for seven out of nine subscales, with Conclusion and Organization being the exceptions. Of note is the fact that, for several outcomes, the fractal properties explain more than twice the variance explained by NW.

Table 2. Regression models for expert rated subscales

	Predictors							
	Model 1		Model 2					
DV	NW	R^2	NW	*H*	$dh(q)$	R^2	*F*
Intro	0.30***	0.10	0.39***	0.27**	0.29***	0.20	7.85***
Body	0.43***	0.22	0.38***	0.18*	0.15	0.27	3.52*
Conc.	0.63***	0.23	0.64***	0.24*	0.13	0.26	2.46
Org.	0.39***	0.15	0.44***	0.16	0.16	0.19	2.45
Coh.	0.17*	0.04	0.23**	0.19*	0.17*	0.10	3.71*
Gram.	0.15*	0.04	0.24***	0.20**	0.33***	0.18	11.30***
Voice	0.30***	0.15	0.34***	0.17*	0.20**	0.22	5.78**
Word	0.22***	0.10	0.31***	0.22***	0.31***	0.27	15.06***
Sent.	0.30***	0.16	0.37***	0.19**	0.23***	0.25	7.79***

Note: ***p < 0.001, **p<0.01, *p<0.05. F test was based on 2 and 127 degrees of freedom.

Discussion

In this study, we investigated how the multiscale characteristics of text production relate to essay quality. In general, we found that the Hurst exponent was a positive predictor of holistic essay quality and analytical scores. Similarly, we found that broader multifractal spectra predicted better quality essays, overall, and on several analytical subscales. The remainder of the discussion is structured as follows: First, we give an overview and basic interpretation of the scaling behavior observed for IKIs during essay production. Second, we speculate on how those interpretations inform patterns of prediction observed with respect to essay quality. Lastly, we offer ideas for potential applications and future research.

Scaling Properties in IKIs

We found that IKIs in this study were characterized by global *H* close to the value typical of random variation (i.e., $H = 0.50$). The observed mean of *H* was surprising given the prevalence with which *H*s indicative of long-range correlation (i.e., $H > 0.5$) have been observed in other tasks (Kello et al., 2010). The result was further surprising because keystrokes observed during typing tasks have been previously characterized as being anti-correlated, where $H < 0.5$ (Wallot & Grabowski, 2013). One possible reason for the difference in results is those authors' use of power spectral density to estimate *H* (labeled α in that study). Simulation work has shown power spectral density underestimates *H* (Delignières et al., 2006). The reason for those differences may be the method used to estimate *H*.

A more likely and substantive reason relates to nature of the tasks used in each study. In Wallot and Grabowski (2013), participants performed one of three relatively simple tasks: type a nursery rhyme from memory; copy text from a page; and generate a novel set of directions from school to home. The latter condition was the closest to essay writing but still differs substantially in complexity and difficulty, factors known affect the scaling properties in basic motor control tasks and complex tasks like steering (e.g., Chen et al., 2001; Likens et al., 2015). Writing a timed essay is arguably more difficult and more complex than giving familiar directions. Perhaps, then, the Hurst exponents we observed in this study reflect those differences in task difficulty. The results concerning $dh(q)$ lends further, albeit tentative, support for that conclusion.

The trend across tasks observed in Wallot and Grabowski (2013) permits cautious speculation about the meaning of spectral widths within the current context. Note that a direct comparison between the widths we observed and those in Wallot and Grabowski is not possible because they used a wavelet form of multifractal analysis, and different methods are known to produce different widths (Ihlen & Vereijken, 2013). In Wallot and Grabowski, the multifractal spectrum width increased as a function of task complexity, with the generative task producing the broadest spectrum. Similar results have also been reported in the motor control and social coordination literatures where an increase in task difficulty has been associated with widening $dh(q)$ (e.g., Davis, Brooks, & Dixon, 2016; Romero, Coey, Beach, & Richardson, 2013). A reasonable conclusion is that the relatively broad spectra we observed reflect the difficulty inherent in writing a timed essay.

Unlike *H*, the multifractal spectrum does not have the same useful interpretive indices concerning long range correlation and randomness. However, a few words are possible concerning why task complexity or task difficulty would affect the width of the multifractal spectrum. The multifractal spectrum provides a summary of scaling behaviors that evolve over time (Ihlen & Vereijken, 2013; Kantelhardt et al., 2002). If the Hurst exponent were sufficient to describe the scaling behavior present in the IKI time series, then one would expect a narrow spectrum – a time-invariant monofractal process. Instead, we observed broad multifractal spectra that are more consistent with interpretation of a time-varying multifractal process. Time-varying scaling behavior is thought to reflect the ongoing dynamics in complex, dynamical systems that range from individual physiological processes to entire human teams (Likens et al., 2014). Time-varying scaling behavior in the

IKIs might reflect changes in cognitive state or changes in strategy that accompany the multiscale coordination involved in writing an essay (e.g., Stephen et al., 2009). That idea is elaborated in the following section in the context of essay quality.

Scaling Properties as Predictors of Essay Quality

We have suggested that changes in scaling behavior may reflect changes in cognitive state or strategy. If so, then a broader multifractal spectrum could reflect flexibility in writing. Multifractal scaling is synonymous with flexibility and adaptability in other contexts (e.g., Collins & De Luca, 1993), and flexibility in the use of cohesive devices (i.e., flexibility in writing) predicts higher quality essays (e.g., Allen et al., 2016; Snow et al., 2015). The implication is: if multifractal scaling reflects flexibility in writing, then wider multifractal spectra may also indicate higher quality essays. The current findings seem to support such reasoning. Results from regression analyses suggest fractal properties positively predict overall essay quality as well as quality on analytical subscales.

Another notable feature of the regression analyses was that $dh(q)$ did not predict the quality of either the Body or Conclusion. As a potential explanation of those results, we refer to our data preparation steps. The time series in our sample were truncated to accommodate participants with short essays. Given the average length of intact series was over three times the length of the truncated series we analyzed, there is a strong possibility the fractal analyses did not equally represent Body and Conclusion aspects of text. If true, then perhaps $dh(q)$ did not adequately capture variability with respect to Body and Conclusion sections. In addition, neither H nor $dh(q)$ predicted the organization subscale. A similar interpretation could be made concerning the length of the time series analyzed with respect to the length of a typical essay in our sample.

Applications and Future Directions

In this study, we have shown for the first time that fractal properties measured while writing an essay predict essay quality. Being the first of its kind, we have interpreted the results cautiously. However, the results are promising and suggest opportunities for future research and applications.

One promising area of research pertains to flexibility and adaptability in writing. As already discussed, multifractal scaling may suggest flexibility and adaptability in writing. If so, then it should be possible to link multifractal characteristics with other aspects of writing flexibility (Allen et al., 2016; Snow et al., 2015). In those studies, flexibility was characterized over several essays; however, if flexibility is important on the timescales of days and weeks, then flexibility should also be important within the context of a single essay. If so, then the fractal properties of keystrokes may also relate to the flexibility at those slower time scales.

Another related area of investigation involves the use of fractal properties in applied settings. The results of the current study, if replicable, could inform applied educational settings such as those involving learning analytics and automated writing evaluation systems. The analyses we have presented here are algorithmically efficient enough to be implemented in real time. Real-time assessment of fractal properties is promising on several fronts. Real-time fractal properties could be monitored by instructors for early signs of writing difficulty and provide faster, targeted feedback. The same notion could apply within automated writing evaluation systems to augment automated feedback systems.

Lastly, the methods we have presented are not limited to the analysis of keystrokes. The use of physiological measurements and various movement registration devices is becoming more common in applied literature on intelligent tutoring systems (D'Mello, Picard, & Graesser, 2007). Fractal analyses have proven beneficial in other settings involving physiological data, primarily because of relationship between fractality and flexibility (e.g., Chen et al., 2001; Ivanov et al., 2001). An open, empirical question is whether fractal analysis of physiological data may reveal flexibility in intentional forms of behavior. In conclusion, we have demonstrated that text production exhibits scaling properties like those observed in other cognitive phenomena. In doing so, we have also supported the idea that language is a complex, dynamical system involving coordination across many nested time scales. Going forward, our goal will be to further articulate time scales relevant to text production.

Acknowledgments

This research is supported in part by the Institute of Education Sciences (R305A120707) and the Office of Naval Research (N00014140343 and N12249156). Any opinions or conclusions expressed are those of the authors and do not represent views of the IES or ONR.

References

Allen, L., Snow, E., & McNamara, D. (2016). The Narrative Waltz: The Role of Flexibility in Writing Proficiency. *Journal of Educational Psychology, 108*, 911-924.

Anastas, J., Stephen, D., & Dixon, J. (2011). The scaling behavior of hand motions reveals self-organization during an executive function task. *Physica A: Statistical Mechanics and its Applications*, *390*(9), 1539-1545.

Beran, J. (1994). *Statistics for long-memory processes.* Boca Raton, FL: Chapman and Hall/CRC Press.

Chen, Y., Ding, M., & Kelso, J. (1997). Long memory processes (1/f α type) in human coordination. *Physical Review Letters*, *79*(22), 4501-4504.

Chen, Y., Ding, M., & Kelso, J. A. (2001). Origins of timing errors in human sensorimotor coordination. *Journal of Motor Behavior*, *33*(1), 3-8.

Collins, J. J., & De Luca, C. J. (1993). Open-loop and closed-loop control of posture: A random-walk analysis of center-of-pressure trajectories. *Experimental Brain Research, 95*, 308-318.

Dale, R., Kello, C., & Schoenemann, P. (2016). Seekng synthesis: The integrative problem in understanding language and its evolution. *Topics in Cognitive Science, 8,* 371-381.

Davis, T. J., Brooks, T. R., & Dixon, J. A. (2016). Multi-scale interactions in interpersonal coordination. *Journal of Sport and Health Science*, *5*(1), 25-34.

Delignières, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for 'short' time series: a re-assessment of classical methods. *Journal of Mathematical Psychology*, *50*(6), 525-544.

D'Mello, S., Picard, R., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*(4).

Eke, A., Herman, P., Kocsis, L., & Kozak, L. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, *23*1), R1-R38.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/*f* noise in human cognition. *Science*, *267*, 1837-1839.

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear Dynamics, Psychology and Life Sciences*, *14*, 265-289.

Ihlen, E., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond $1/f^{\alpha}$ fluctuation. *Journal of Experimental Psychology: General*, *139*, 436-463.

Ihlen, E., & Vereijken, B. (2013). Multifractal formalisms of human behavior. *Human Movement Science*, *32*, 633-651.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, *316*(1), 87-114.

Kelty-Stephen, D. G., & Wallot, S. (in press). Multifractality versus (mono)fractality as evidence of nonlinear interactions across time scales: Disentangling the belief in nonlinearity from the diagnosis of nonlinearity in empirical data. *Ecological Psychology.*

Likens, A. D., Amazeen, P. G., Stevens, R., Galloway, T., & Gorman, J. C. (2014). Neural signatures of team coordination are revealed by multifractal analysis. *Social Neuroscience*, *9*(3), 219-234.

Likens, A., Fine, J., Amazeen, E., & Amazeen, P. (2015). Experimental control of scaling behavior: what is not fractal? *Experimental Brain Research*, *233*, 2813-2821.

Mandelbrot, B., & Van Ness, J. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM review*, *10*(4), 422-437.

McNamara, D., Crossley, S., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, *45*, 499-515.

McNamara, D., Crossley, S., Roscoe, R., Allen, L., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*, 35-59.

Nonaka, T., & Bril, B. (2014). Fractal dynamics in dexterous tool use: The case of hammering behavior of bead craftsmen. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 218.

Pinet, S., Ziegler, J., & Alario, F. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin & review*, *23*, 1898-1906.

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, *49*(2), 1685 - 1689.

Rączaszek-Leonardi, J. (2010). Multiple time-scales of language dynamics: An example from psycholinguistics. *Ecological Psychology*, *22*(4), 269-285.

Rączaszek-Leonardi, J., & Kelso, J. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology*, *26*, 193-207.

Romero, V., Coey, C. A., Beach, A., & Richardson, M. J. (2013). Effects of Target Size and Symmetry on the Structure of Variability in Precision Aiming. In *CogSci*.

Snow, E. L., Allen, L. K., Jacovina, M. E., Crossley, S. A., Perret, C. A., & McNamara, D. S. (2016). Keys to Detecting Writing Flexibility Over Time: Entropy and Natural Language Processing. *Journal of Learning Analytics*, *2*(3), 40-54.

Stephen, D. G., & Anastas, J. (2011). Fractal fluctuations in gaze speed visual search. *Attention, Perception, & Psychophysics*, *73*(3), 666-677.

Stephen, D. G., Arzamarski, R., & Michaels, C. F. (2010). The role of fractality in perceptual learning: Exploration in dynamic touch. *Journal of Experimental Psychology. Human perception and performance*, *36*(5), 1161.

Stephen, D. G., Broncoddo, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition*, *37*(8), 1132-1149.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*, 331.

Wagenmakers, E. J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/fα noise in human cognition. *Psychonomic Bulletin & Review*, *11*, 579-615.

Wallot, S., & Grabowski, J. (2013). Typewriting Dynamics: What Distinguishes Simple From Complex Writing Tasks?. *Ecological Psychology*, *25*(3), 267-280.

Wallot, S., Hollis, G., & van Rooij, M. (2013). Connected text reading and differences in text reading fluency in adult readers. *PloS One*, *8*(8), e71914.

Wallot, S., & Van Orden, G. (2011). Toward a lifespan metric of reading fluency. *International Journal of Bifurcation and Chaos*, *21*(04), 1173-1192.

Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia, 62*, 100-119.