Journal of the American Society of Questioned Document Examiners, Inc.
Volume 21, Number 1, June 2018

Kinematic Validation of FDE Determinations about Writership in Handwriting Examination: A Preliminary Study

Michael Caligiuri[1], Linton Mohammed[2], Brenda Lanners[3], and Gina Hunter[3]

[1]Department of Psychiatry, University of California San Diego, 9500 Gilman Drive (0603), La Jolla, CA 92093, USA

[2]Forensic Science Consultants, Inc., 433 Airport Blvd., Suite 406, Burlingame, CA 94010, USA

[3]San Diego Sheriff's Crime Lab, 5590 Overland Ave, San Diego, CA 92123, USA

As with many of the forensic disciplines that rely on feature-comparison methods, there is no "gold standard" against which to test accuracy of handwriting examination. This study examined differences in kinematic features between pairs of handwriting exemplars judged to be from the same writer and compared them with differences in kinematic features for pairs of handwriting exemplars judged to be from different writers. We hypothesized that differences in kinematic features between pairs of handwriting exemplars judged to be from the same writer would be nonsignificant; whereas differences in kinematic features for pairs of handwriting exemplars judged to be from different writers would be statistically significant. Cursive, script and block print handwriting samples were obtained from 37 writers who were asked to write a single word ten times each. High resolution (600 ppi) scanned copies of the original ink and paper samples were submitted to four experienced forensic document examiners (FDEs) for writership determinations. Each score sheet included 5 known (K) handwritten samples and two questioned samples (Q1 and Q2). FDEs were asked to rate the evidence in support for the proposition that the Q samples were written by the K writer using a 4-point scale (ranging from limited or weak support to very strong support for the proposition). Kinematic difference scores derived from dynamic analysis of the handwritten strokes were converted to absolute standardized z-scores with larger z-score reflecting greater differences between K and Q for a given kinematic feature. Findings revealed that several kinematic handwriting features were significantly associated with accurate FDE opinions of acceptance and rejection of the proposition. Significant features included pen pressure, stroke velocity, and straightness variability. Correlational analyses revealed strong associations between specific dynamically recorded stroke features and FDE judgments of writership; particularly for pen pressure and straightness. Results support the use of an independent quantitative measure of feature comparison as a tool for evaluating the foundational validity of subjective feature comparison methods experts use when reaching conclusions about writership.

Introduction

Forensic handwriting examiners rely on feature-comparison methods to determine whether an evidentiary sample is or is not associated with a potential "source" sample (e.g., from an individual), based on the presence of patterns, impressions, or features appearing in both samples. Using standard comparison methods, the available literature from signature and handwriting authentication studies report accuracy rates ranging from 86% to 96% for trained document examiners [1-7]. While these findings demonstrate that expert examiners are proficient and perhaps reliable (at least in a laboratory setting), they cannot address the validity of their opinions, as there has been no independent "gold standard" or litmus test against which to validate expert opinion.

Challenges to the admissibility of handwriting expert opinions generally derive from the lack of empirical validation of conventional methods.

In handwriting comparisons, the forensic document examiner constitutes a significant part of the measurement instrument. Upon observing the features in the samples, the examiner gauges their importance using various scales ranging from a 3-point scale (same source, inconclusive, or different source) to the more elaborate Scientific Working Group for Documents (SWGDOC) [8] 9-point classification scheme for the strength of the examiner's opinion. Unfortunately, the available science on validation of an examiner's opinion is limited to studies of error rates. Empirical studies to validate feature comparisons using quantitative independent measures have not been conducted. Saks and Koehler [9] observed that while proficiency tests represent a step in the right direction toward validation, questionable generalizability to actual casework and infrequent peer review has been a barrier to admissibility of expert opinion under *Daubert.*

When expert handwriting examiners follow accepted best practices, they compare specific features between questioned and known samples to estimate a probability that the samples were written by a single writer. While such comparative methods are largely subjective, research supporting the foundational validity of this approach, that is, validity based on scientific principles or theories is lacking.

Laboratory research on handwriting kinematics published over the past 35 years has contributed to the development of a reliable quantitative method for extracting specific features from handwriting samples [10-16]. The dynamic methodology yields numerous independent features characterizing the spatial and geometric characteristics of pen strokes (often referred to as the dynamic approach). In one noteworthy study based on dynamic handwriting features, Ostrum and Tanaka [12] asked FDEs to examine handwritten paper samples and make judgments about writing speed, stroke angle, pressure variation, and other features commonly examined in FDE casework. These judgments were then compared with independent dynamic data of the same samples obtained from digital recordings. Although the study did not subject these data to rigorous statistical procedures (due to a small sample size), the study revealed remarkable agreement between FDE judgments and dynamic analyses, particularly for writing speed and pressure variation. The authors concluded that analysis of dynamic features is an effective method for the validation of FDEs' abilities to extract characteristic from the static written record.

The present study is one of few to cross-validate examiner determinations of writership using an established dynamic approach to extract specific geometric, spatial, and kinematic features of handwriting and hand printing. This study examined differences in kinematic features between pairs of handwriting exemplars judged to be from the same writer and compared them with differences in kinematic features for pairs of handwriting exemplars judged to be from different writers. We hypothesized that differences in kinematic features between pairs of handwriting exemplars judged to be from the same writer would be non-significant; whereas differences in kinematic features for pairs of handwriting exemplars judged to be from different writers would be statistically significant. A secondary aim was to test whether specific handwriting kinematic features associated with accurate FDE opinions of writership differed across three styles of handwriting.

Methods

Handwriting samples were obtained from 37 writers recruited from the employees of the San Diego Sheriff's Crime Laboratory. Writers had a mean age of 38.71 (sd=8.93) years. The majority were female (n=27) and right-handed (n=34).

Study subjects were asked to write a single word "Alabama" ten times using each of three writing styles: cursive, script, and block. We chose a single word for this pilot study to reduce variability in kinematic features across writers making the task more challenging to examiners particularly for the printed samples. The word "Alabama" was selected because it contains multiple instances of a single character that is likely to be produced with minimal variation in stroke features by the same writer. A long-term goal of this research was to examine intra-writer variation in stroke features and we reasoned that the word "Alabama" would provide sufficient number of duplicated features to reliably estimate within- and between-word variability without potential confounds associated with writer fatigue.

Writers were instructed to write the target word with their preferred hand once per trial for ten trials using an inking pen (stylus) on unlined

8.5 x 11″ pieces of white paper beginning at the upper left and descending in column format. Separate sheets of paper were used for each of the writing styles. Style order was randomized across subjects. Writers returned to the laboratory for a second handwriting session three weeks later to provide handwriting samples. Handwriting exemplars from the second visit were paired with those from the same writer for the first visit to create pairs where the questioned sample is a close match to the known samples. By including known and questioned exemplars from the same writer but different days, the task of determining writership becomes more difficult for the examiner and more closely resembles actual casework.

We followed standard published procedures for digitizing the samples and extracting kinematic features from handwriting samples [13-14,17]. The procedures involved the use of an inking pen with a Wacom[1] Intuos Pro UD 9 x 12 digitizing tablet (30 cm x 22.5 cm, sampling rate 120 Hz, RMS accuracy 0.01 cm) attached to a notebook computer running MovAlyzeR®[2]. MovAlyzeR® software was chosen for this project because it is capable of reliably extracting multiple kinematic and geometric variables from each pen stroke. The software allows precise extraction of pen movement in the time (x), amplitude (y), and pressure (z) dimensions.

A Wacom inking stylus was used for all of the handwriting samples collected. This stylus was chosen as the feel is similar to that of a ball-point pen with which the subjects will be familiar and presented a naturalistic writing condition. Each sheet of unlined paper was positioned over the digitizing pad in a fixed position; however, the writer was permitted to reposition the tablet to achieve a comfortable writing position while seated.

Multiple kinematic parameters along with pen pressure were extracted from each vertical and horizontal pen stroke using MovAlyzeR® software. Table 1 shows the stroke parameters examined in this study along with their operational definitions. Vertical stroke movements were segmented using the local minima of the absolute velocity time curve, that is, the time points when vertical pen movement direction changes [18]. Features associated with upstrokes were examined separately from downstrokes. This is especially important for slant, as vertical upstrokes produced with a right slant would be characterized by stroke angles typically between 20º - 30º; whereas vertical downstrokes are commonly produced with minimal tilt and would be characterized by stroke angles near 180º degrees.

High resolution (600 ppi) scanned copies of the original ink and paper samples were arranged on a single sheet of paper and submitted to one of

Table 1. Handwriting stroke parameters used in this study and their operational definitions[3].

Stroke Parameter	Definition
Duration	Time interval (sec) between the first and last samples in a stroke
Vertical Amplitude	Vertical vector difference between beginning and end of a stroke (cm).
Horizontal Amplitude	Horizontal vector difference between begin and end of a stroke
Peak Velocity	First derivative of vertical displacement (cm/s)
Peak Acceleration	Second derivative of vertical displacement (in cm/s^2)
Straightness Error	Straightness error or the normalized standard deviation from a straight baseline is the RMS difference from the straight line fitted through the stroke divided by the stroke distance (StraightErr = (1/length) * sqrt (Sum (y(t) - yd(t))**2 / N)). A perfectly straight stroke will have straightness error of 0. It is orientation free.
Slant	The angle or inclination of the axes of letters relative to the perpendicular to the baseline of the writing (in radians).
Pen Down Duration	Ratio of pen up duration to total duration. The value ranges from 0 to 1. 0 corresponds to no pen down and 1 corresponds to all pen down and the values in between show the ratio.
Loop Surface	Surface or the area of the loop enclosed by the previous and present stroke in cm**2. The surface is not normalized. If the the crossing does not occur within the previous stroke, although a loop has been formed, the loop area will be zero.
Pen Pressure	Relative axial pressure on the pen tip when the pen is on the paper (ranging from 0-2047).

[1] http://www.wacom.com/en-us. Last accessed on February 23, 2018
[2] www.neuroscriptsoftware.com. Last accessed on February 23, 2018
[3] http://www.neuroscript.net/help/viewingtrials.html. Last accessed on February 23, 2018

the co-authors (L.M.) to identify 10 pairs of words that "looked very similar". This was done for each of the three writing styles. The purpose of this exercise was to increase the difficulty of the FDE opinion survey by eliminating from the pool of exemplars those handwriting samples which appeared uncommon and easy to exclude as written by another writer. From each of the 10 pairs (20 writers), one of the pair was assigned as a known (K) sample, while the other was assigned as a questioned (Q1) sample. A second questioned sample (Q2) came from the second set of samples written by one of the two writers (either K or Q1) obtained three weeks after the first set. Thus, of the two questioned samples, one was actually from the same writer as K (acquired from a second visit), while the other questioned sample was selected to closely resemble the known samples from the larger pool of 37 writers. Each opinion survey included 10 triads (K, Q1, Q2) of the word "Alabama" written in three styles for a total of 60 opinions (10 triads x 2 questioned samples x 3 handwriting styles). Figure 1 shows examples of an FDE opinion scoresheet for the three handwriting styles. Sample pairings were created so that there were 30 word pairs written by the same writer and 30 pairs written by different writers (10 each for three writing styles).

FDEs were provided with the following instructions: "Shown on each page are two sets of 5 samples of a single word from a known writer along with two questioned samples. Examine these samples and "score" each of the two questioned samples as likely written by the known writer. First, rate the evidence in support for the proposition that Q1 is the same writer as the Known writer. Then score Q2 for the proposition that Q2 is the same writer as the Known writer. Provide two scores (Q1 and Q2) for each set using the following scale: 1 = limited or weak support for proposition; 2 = moderate support for proposition; 3 = strong support for proposition; and 4 = very strong support for proposition.

Opinion surveys were sent to five certified FDEs who expressed interest in participating. Four FDEs completed and returned the surveys. These FDEs had a mean of 26.5 years of experience with a range of 20 to 40 years. Their mean judgment ratings for each of the 60 sample pairs were entered into a statistical database along with corresponding kinematic data for each K-Q1 and K-Q2 pair. Based on the mean ratings FDE opinions were classified into four categories: (1) true accept for the proposition that Q is the same writer as K when Q is the same writer as K (mean opinion score ≥ 2.25) ; (2) true reject for the proposition that Q is the same writer as K when Q is a different writer as K (mean opinion score < 1.50); (3) false acceptance or judging that K and Q samples are from the same writer when in fact they are not; and (4) false rejection or judging that K and Q samples are from different writers when in fact they are from the same writer. In the present study, we present findings only for the true accept or true reject opinions because sample sizes for categories 3 and 4 were insufficient for statistical analyses.

Kinematic difference scores (K-Q1 and K-Q2) were converted to absolute standardized Z-scores using the formula $Z = (K\text{-}Qn)/Ksd$; where K is the mean kinematic value across all strokes for a single known sample, Q is the mean kinematic value across all strokes for a single question-d sample (n=1 or 2), and Ksd is the standard deviation of the first five known samples obtained from a given writer during the first of two handwriting sessions (also shown on each score sheet for the K samples). Absolute Z-score values were used as there was no *a priori* reason to expect K-Q differences to have directional effects. Larger Z-scores are associated with greater standardized differences between K and Q for a given kinematic feature. Standardized scores allow direct comparison across multiple kinematic variables having different units of measure.

For each K-Q pair, the data available for statistical analyses consisted of the mean FDE rating and a Z-score, reflecting the kinematic difference between known and questioned samples for each of 15 kinematic variables for up and downstrokes. For variables satisfying assumptions for parametric statistics, we used independent group t-tests to test whether the kinematic standardized Z-scores differed between True Accept and True Reject determinations. Variables not satisfying assumptions for parametric statistics were examined using nonparametric Mann-Whitney statistics. We also examined the relationships between FDE confidence rating and standardized Z-score using Pearson correlation. Difference tests and correlation coefficients with alpha $\leq$ 0.05 were considered statistically significant.

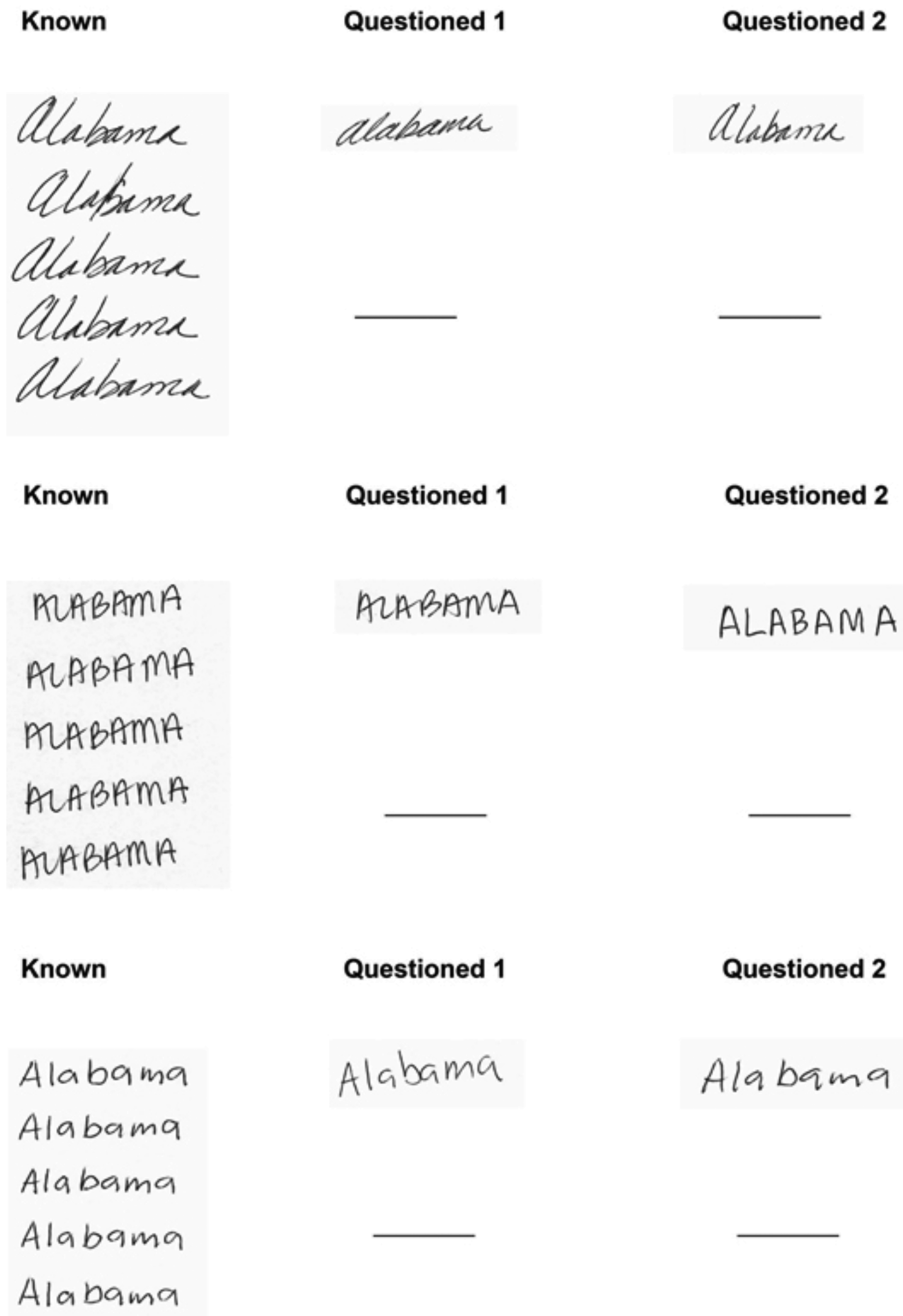

Figure 1. Sample scoresheets showing 5 known specimens (left) and two questioned specimens for cursive (top), block (middle), and script (bottom) handwriting.

Results

FDE Opinion Scores and Accuracy

Based on FDE scores and knowledge of ground truth, we calculated the accuracy rates for sample pairs. Of the 30 pairs from the same writer, FDEs scored strong or very strong support for the same writer 80% of the time. Of the 30 pairs from different writers, FDEs scored strong or very strong support for different writers 97% of the time. When examined by handwriting style, FDEs scored strong or very strong support for the same writer with accuracy rates of 80%, 90%, and

Table 2. Absolute Z-scores for differences in kinematic parameters between K and Q samples for the true accept and true reject classifications for cursive, block, and script style handwriting. Shown are those parameters with p-values ≤ 0.10 for difference tests between accept and reject opinions.

Stroke Direction	Parameter	Accept	Reject	Statistic
Cursive				
Up	Straightness Error	0.23 (0.12)	0.81 (0.76)	t=2.10 ($p \leq 0.05$)
Up	Peak Velocity	0.23 (0.14)	0.48 (0.31)	t=1.99 ($p < 0.10$)
Up	Pen Pressure	0.55 (0.32)	1.92 (2.08)	t=1.83 ($p < 0.10$)
Down	Peak Velocity	0.19 (0.15)	0.51 (0.22)	t=3.35 ($p \leq 0.01$)
Down	Pen Pressure	0.75 (0.50)	2.13 (1.76)	t=2.13 ($p \leq 0.05$)
Down	Slant	0.14 (0.11)	0.37 (0.31)	t=1.93 ($p < 0.10$)
Block Print				
Up	Pen Pressure	0.16 (0.10)	0.46 (0.23)	t=3.24 ($p > 0.005$)
Down	Pen Down Duration	0.13 (0.09)	0.32(0.26)	t=1.82 ($p < 0.10$)
Script Print				
Down	Peak Acceleration	0.36 (0.32)	0.84 (0.63)	t=2.13 ($p < 0.05$)
Down	Pen Pressure	0.26 (0.24)	0.65 (0.61)	t=1.75 ($p > 0.10$)

70% based on ground truth for cursive, block and script handwriting respectively. FDEs scored limited or weak support for the single writer proposition with accuracy rates of 100%, 90%, and 100% based on ground truth for cursive, block and script handwriting respectively.

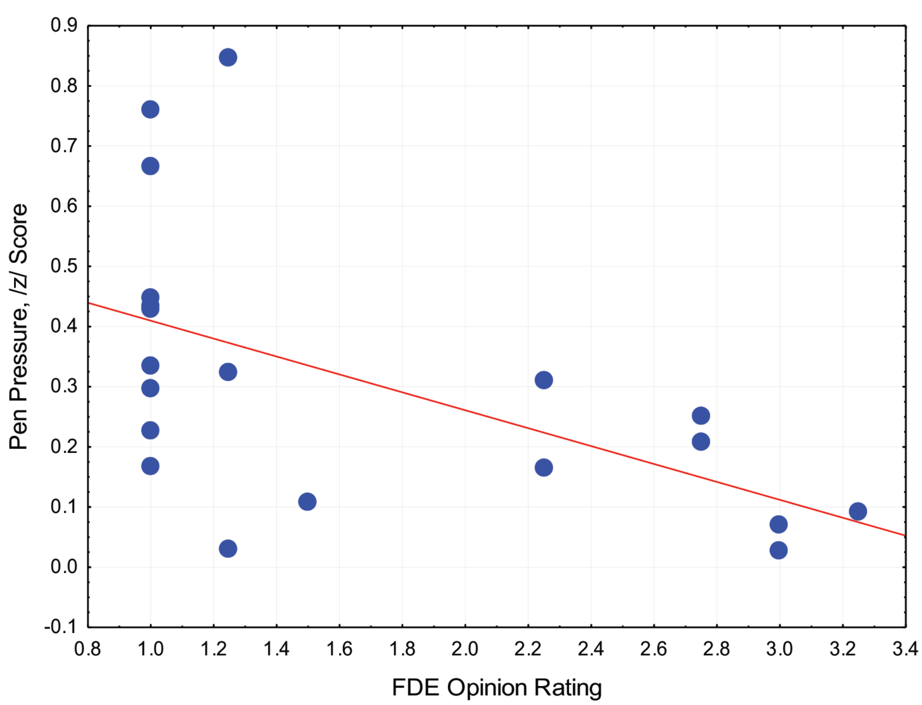

Figure 2. Scatterplot of the relationship between standardized difference score (z) and FDE opinion rating for pen pressure from cursive downstrokes. Shown in the graph is the line of best fit derived from a linear regression model.

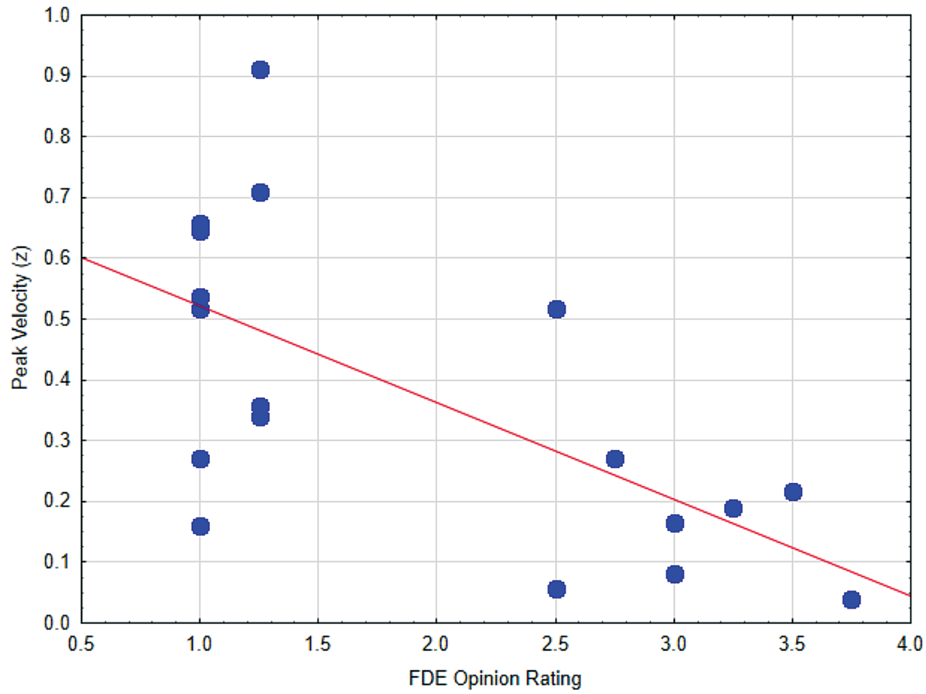

Figure 3. Scatterplot of the relationship between standardized difference score (z) and FDE opinion rating for peak velocity from cursive downstrokes. Shown in the graph is the line of best fit derived from a linear regression model.

Kinematic Features Associated with FDE Opinions

Table 2 shows the absolute Z-scores for differences in kinematic parameters between K and Q samples for the true accept and true reject classifications for cursive, block, and script style handwriting. The table includes only those parameters with p-values ≤ 0.10 for difference tests between accept and reject opinions. Consistent with our hypotheses, mean Z-scores for true accept (i.e. same writer) were lower than that for true reject (i.e. different writer) determinations from experienced FDEs. Stroke features with statistically significant ($p \leq 0.05$) standardized difference scores between FDE accept and reject opinions included: straightness error (upstrokes), peak velocity (downstrokes), and pen pressure (downstrokes) for cursive handwriting; pen pressure (upstrokes) for block print; and peak acceleration (downstrokes) for script print. Trends for statistical significance were observed for peak velocity (upstrokes), pen pressure (upstrokes), and slant (downstrokes) for cursive writing; pen down duration (downstrokes) for block print; and pen pressure (downstrokes) for script print. Statistical tests for other kinematic features such as stroke duration, vertical and horizontal amplitude, and loop surface were nonsignificant.

Correlational analyses were performed to examine relationships between the kinematic Z-scores and FDE judgments of writership. Strong associations were found between pen pressure ($r=-0.65$; $p<0.01$) and peak velocity ($r=-0.65$; $p<0.01$) for cursive downstrokes and FDE opinion scores. These results are shown in Figures 2 and 3. Negative associations indicate that as strength of FDE opinion that two samples were written by the same writer increases, the standardized kinematic difference score for the word pair decreases.

Discussion

This study was the first step in a larger effort to examine the foundational validity of expert opinion using methods derived from studies of handwriting motor control. We identified several kinematic handwriting features that were statistically associated with accurate FDE opinions of acceptance and rejection of the propositions that two handwritten samples were from the same writer. Our results support the hypotheses that differences in kinematic features between pairs of handwriting exemplars judged to be from the same writer are smaller than differences in kinematic features for pairs of handwriting exemplars judged to be from different writers. Features associated with accurate FDE judgments of writership included pen pressure, stroke velocity, stroke angle, straightness variability, and loop surface. Statistically significant differences in stroke kinematics between matched and mismatched pairs were present across the three writing styles; although more kinematic differences were observed for cursive handwriting than script or block printing. Statistically significant correlations were observed for upstroke straightness error for cursive writing, downstroke peak velocity for cursive writing, and upstroke pen pressure for script printing.

The present findings are consistent with a previous study by Ostrum and Tanaka [12]. The authors found remarkable agreement between FDE judgments and dynamic analyses of several dynamic features including amplitude, velocity and applied pen force at multiple segmentation points along the time-series. Using different methods and means for measuring dynamic handwriting features, the present study also identified a strong statistical relationship between such features as pen pressure and stroke velocity and FDE judgments of writership.

President's Council of Advisors on Science and Technology (PCAST), Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods [19] recommended more empirical research to assess the foundational validity in the forensic sciences. The PCAST report included several recommendations to improve measurement validity in the forensic sciences including white-box studies to understand methods used in feature comparison and to develop technology for more objective measures. White-box studies are designed to evaluate the relationships between feature characteristics and examiner decisions. The results of the present study suggest that examiners' putative internalized processes likely involve the recognition and weighting of specific handwriting features such as stroke slant, pen pressure, straightness variability, and loop area; although such recognition is probably not overt to the examiner. Though not likely to be directly observed from careful examination of handwriting, dynamic features such as stroke velocity and acceleration may have indirectly influenced FDE opinion of writership. Interestingly many of the parameters found to be significantly associated with FDE opinions of writership are ones that can be observed from static handwriting images. These features include pen pressure, straightness error, stroke angle and loop surface area.

The present study provides preliminary empirical support for current methods FDEs use to express opinion. PCAST [19] defines foundational validity as the "*scientific* standard corresponding to the legal standard of evidence being based on reliable principles and methods." (p. 43). The present study provides some empirical support suggesting that FDE writership opinions may be informed by scientific principles of handwriting motor control which posits that handwriting is represented by a flexible generalized motor program containing the timing, sequence, and spatial patterns of pen strokes available to the habitual writer as a single action sequence to form letters and words.

The motor control theory of handwriting [20-22] provides an empirically based justification for why features such as stroke slant, amplitude, or pen pressure vary across individuals and to a lesser extent, within an individual over time. Variability in these patterns and features is an attribute of a flexible generalized motor program that informs the FDE about a writer's range of natural variability. These principles were supported by the present findings showing significantly greater differences for several kinematic features when calculated from samples written by different individuals than samples written by a single writer.

The present study has limitations and should be considered preliminary. Limitations include: the use of single-word pairs as the only material available to reach a writership determination; the small number of exemplars (n=20 pairs for each writing style); and our decision to rate confidence of writership determinations using a 4-point scale of certainty. The use of single word pairings for a laboratory-based writership matching experiment was desirable in the present study to reduce kinematic variability; however as a performance measure it is inadequate. With regard to the sample sizes ranging from 7-10 exemplar pairs (per handwriting style) accurately judged as either from the same or different writers, it is possible that with larger sample sizes, more kinematic stroke feature difference scores would reach statistical significance. Even with our modest sample size, important handwriting features were found to be significantly associated with FDE writership opinions. The selection by a study co-investigator of the word pairs that looked very similar for inclusion in the final survey was entirely subjective and based on personal experience. This may have introduced bias which could be managed in future studies by randomly matching pairs of single words. While random pairing would increase generalizability of the findings, this requires a greater number of samples that collected in the present study. Larger scale follow-up studies currently underway will deploy automated systems based on visual feature matching rather than human decision-making to identify word and phrase samples that are similar for the purpose of validating examiner writership judgments. Finally, caution must be exercised not to over-interpretation of the FDE accuracy rates reported in this study, as the task is not representative of typical cases found in practice, nor would the prudent examiner likely consider single-word pairs to be sufficient for a reliable examination.

Currently there is no consensus within the discipline on whether to apply a 4-, 5-, or 9-point classification scheme to rate FDE confidence in writership opinions. Reliability studies have not been conducted. While a 4-point classification scale would be inappropriate for judging the strength of two mutually exclusive propositions within the framework of evaluative reporting [23-24], it was important in this preliminary study to minimize the number of "inconclusive" opinions that are common when using more complex scales (e.g. SWGDOC [8]).

Future research is needed to replicate these findings. While increasing the number of participating FDEs is not likely to alter the findings (as the design is likely to use the mean FDE opinion score across examiners and the number of FDEs may not impact the mean opinion score), designing a study with more pairings and word or phrase samples extracted from written paragraphs (such as the London Letter) may very well reveal additional features associated with FDE writership opinions. Such studies could be designed to capture FDE writership opinions for two mutually exclusive propositions.

In conclusion, the results from this preliminary study support the use of independent quantitative measures of feature comparison (i.e. kinematic difference scores) as a tool for evaluating the foundational validity of subjective feature comparison methods experts use when reaching conclusions about writership. Using independent measures, the present study offers reasonable validation of FDE opinions of writership.

Acknowledgments

The authors acknowledge the staff and management of the San Diego Sheriff's Crime Lab for providing the writing samples used in this study.

References

1. Kam, M., Wetstein, J., and Conn, R. (1994) Proficiency of professional document examiners in writer identification. *J Forensic Sciences*, *39*: 5-14.

2. Kam, M., Fielding, G., and Conn, R. (1997) Writer identification by professional document examiners. *J Forensic Sciences*, 42: 778-786.

3. Kam, M. and Lin, E. (2003) Writer identification using hand-printed and non-hand-printed questioned documents. *J Forensic Sciences*, 2003; 48: 1391-1395.

4. Kam, M., Abichandani, O., and Hewett, T. (2015) Simulation detection in handwritten documents by forensic document examiners. *J Forensic Sciences* 60: 936-941.

5. Galbraith, O., Galbraith, C.S., and Galbraith, N.G. (1995) The principle of the Drunkards search as a proxy for scientific analysis: The misuse of handwriting test data in a law journal article. *Int J Forensic Doc Examiners*, 1:7-17.

6. Sita, J., Found, B., and Rogers, D.K. (2002). Forensic handwriting examiners' expertise for signature comparison. *J Forensic Sciences*, 47: 1-8.

7. Bird, C., Found, B., and Rogers, D. (2010) Forensic document examiners' skill in distinguishing between natural and disguised handwriting behaviors. *J Forensic Sciences*, 55: 1291-1295.

8. SWGDOC (2016) https://www.nist.gov/sites/default/files/documents/2016/10/26/swgdoc_standard_terminology_for_expressing_conclusions_of_forensic_document_examiners_150114.pdf (Last accessed on February 23, 2018)

8. Saks, M.J. and Koehler, J.J. (2005). The coming paradigm shift in forensic identification science. *Science* 309: 892-895

10. van Galen, G. (1984). Structural complexity of motor patterns: A study on reaction times of handwritten letters. *Psychological Research*, 46:49-57

11. van Gemmert , A. and van Galen, G. (1996). Dynamical features of mimicking another person's writing and signature. In: M.L. Simner, C.G. Leedham, & A.J.W.M. Thomassen (Eds.), Handwriting and Drawing Research: Basic and Applied Issues: 459-471. Amsterdam: IOP Press.

1. Ostrum, B. and Tanaka, T. (2006) Another look at handwriting movement. *J American Society of Questioned Document Examiners*. 9:57-67.

12. Mohammed, L., Found, B., Caligiuri, M., and Rogers, D. (2011). The Dynamic Character of Disguise Behavior for Text-Based, Mixed, and Stylized Signatures. *J Forensic Sciences*, Vol. 56; S136-141.

13. Caligiuri, M.P. and Mohammed, L.A. (2012) The Neuroscience of Handwriting: Applications for Forensic Document Examination. CRC Press, Boca Raton; 247 pages

14. Caligiuri, M.P., Mohammed, L.A., Found, B. and Rogers, D. (2012) Nonadherence to the Isochrony Principle in Forged Signatures. *Forensic Science International*, 223: 228-232

15. Mohammed, L., Found, B., Caligiuri, M., and Rogers, D. (2015). Dynamic Characteristics of Signatures: Effects of Writer Style on Genuine and Simulated Signatures. *J Forensic Sciences*, 60: 89-94.

16. Caligiuri, M.P., Chi, K., and Landy, K.M. (2014). Kinematic Analyses of Signature Writing in Healthy Aging. *J Forensic Sciences*, 59:1020-4

17. Teulings, H.-L. and Schomaker, L. (1993). Invariant properties between stroke features in handwriting. *Acta Psychol*. 82: 69–88

18. Report on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods on September 20, 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (Last accessed on February 23, 2018)

19. Plamondon, R., and Djioua, M. (2006). A multi-level representation paradigm for handwriting stroke generation. *Human Movement Science*, 25:586-607.

20. Thomassen, A.J.W.M., and Van Galen, G.P. (1992). Handwriting as a motor task: Experimentation, modeling and simulation. In Approaches to the Study of Motor control and Learning Ed. J.J. Summers pp 113-144. Amsterdam: North Holland.

21. Found, B. (1997). The forensic analysis of behavioral artefacts: Investigations of theoretical and analytical approaches to handwriting identification. Unpublished doctoral thesis, La Trobe University.

22. Adam, C. (2016) Forensic Evidence in Court: Evaluation and Scientific Opinion. John Wiley and Sons, Ltd, West Sussex, UK.

23. Robertson, B., Vignaux, G.A., and Berger, C.E.H. (2016) Interpreting Evidence: Evaluating Forensic Science in the Courtroom. 2nd Ed. John Wiley and Sons, Ltd, West Sussex, UK.