The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	Cognitive Human Factors and Forensic Document Examiner Methods and Procedures
Author(s):	Mara L. Merlino, Chandler Al Namer, Taleb Al Namer, La'Quida Smith, Veronica Blas Dahir, Charles Edwards, Derek L. Hammond
Document Number:	254581
Date Received:	March 2020
Award Number:	2015-DN-BX-K069

Cognitive Human Factors and Forensic Document Examiner Methods and Procedures

Final Summary Overview
NIJ Award Number 2015-DN-BX-K069

Principal Investigator: Mara L. Merlino
Research Assistants: Chandler Al Namer, Taleb Al Namer, La'Quida Smith
Kentucky State University
Frankfort, Kentucky 40601

Subaward Principal Investigator: Veronica Blas Dahir
Research Assistants: Charles Edwards
University of Nevada, Reno
Reno, Nevada 89557

Expert Consultant: Derek L. Hammond
U.S. Army Criminal Investigation Laboratory
Forest Park, Georgia

Acknowledgements: Bryan J. Found, Victoria Police Forensic Services Department
Adrian Dyer, Royal Melbourne Institute of Technology
Kentucky State University: Piarre Easley, Robert Olson
University of Nevada, Reno: Mauricio Alvarez, J. Guillermo Villalobos, Denise Schaar Buis, Emily Wood, Chris Swinger, Chris Sanchez, Katherine Caufield

Submitted to the U.S. Department of Justice, Office of Justice Programs
National Institute of Justice, 810 Seventh Street N.W., Washington, DC 20531
October 3, 2019

AWARDEE: Kentucky State University

Cognitive Human Factors and Forensic Document Examiner Methods and Procedures

This project was supported by Award No. (NIJ Award # 2015-DN-BX-K069), awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

Project Goals and Objectives

The purpose of this research is to address the needs identified by DOJ, NIJ, NSF, NIST, and other organizations by employing multi-disciplinary research for improved understanding of the production of science in service to the legal system and national security interests. Interdisciplinary research encompassing expertise from forensic practice, social and cognitive psychology, vision science, and other areas, is needed to establish the basis and extent of expertise, to develop rigorous protocols and measures, and to establish education and training programs that consistently and comprehensively address the knowledge and skills required to establish expertise in forensic fields.

The proposed international multidisciplinary research program will extend our previous work (NIJ Award No. 2010-DN-BX-K271), by supporting ongoing research which empirically explores the reliability, measurement validity, and accuracy of established FDE procedures. The series of three experiments in this project adds to this body of knowledge by focusing on additional questions about the FDE examination and decision making process, and the nature and psychometric properties of the opinion continuums. The project supported ongoing teaching and training opportunities for students from a historically black college/university (HBCU), as well as students from collaborating institutions.

Our overall research goal is to investigate the influence of possible sources of cognitive bias in the methods and procedures of forensic document examination. Specifically, addressed the following **research questions:**

- *What is the relationship between the context established by presentation order of questioned and known writing and the examination process?*
- *How do examiners apply the currently-used bipolar continuum of certainty (Elimination through Identification with a center position of Inconclusive) when expressing their opinions about the authorship of questioned writings?*
- *How much writing constitutes "sufficient" information upon which to base an opinion?*

We addressed these specific **research goals**:

- Goal 1: Investigate the relationship among human factors such as visual context, semantic content, attentional resources, salience, bottom-up/top-down processing, perception, and feature matching in FDE decision making, and examine the interaction between these factors and task performance in samples of professional document examiners and lay participants.
- Goal 2: Investigate the measurement properties of the nine-point opinion continuum and the utility of fuzzy set theory in quantifying position values along the continuum.

- Goal 3: Investigate the relationship between the amount of handwriting available for examination and the accuracy and certainty of participants' opinions.

Methods

Study design. The study employed three phases and several methodologies. Phase 1 was a multimodal (Internet/phone) survey designed to gather information about the experience, education, and credentials of our participants, and their views of the strengths and weaknesses of forensic education. Phase 2 comprised three experimental eye tracking protocols (i.e., the signature presentation order protocols for experiment 1 and the cumulative information protocol for experiment 3). The use of the opinion measure (experiment 2) was incorporated within the other eye tracking protocols. Phase 3 was an open-ended, qualitative interview with the FDE and Lay participants in which we will elicit verbal descriptions about the participants' decision making processes for a subset of the signatures from the eye tracking protocols. All procedures were double-blind and conducted under controlled conditions.

Participants and Recruiting. Eighty-five government lab-affiliated and independent examiners participated. Of these, 14 (16.5%) were Australian, 13 (15.3%) were Canadian, and the remaining 58 (68.2%) were U.S. examiners. Twenty-eight lay participants were recruited by a local employment agency. We experienced considerable difficulty recruiting lay participants and did not reach our target of 80. Four of the 28 Lay participants did not complete the entire study, although the data they did provide is included in our analyses. This impacted the power and complexity of the analyses we were perform when comparing the performance of FDE and Lay participants.

We recruited participants using a modified snowball technique in which the research team attended professional meetings to present information about the project and to personally invite attendees to participate. Participants were accepted on a first come, first-served basis. Respondents were free of visual impairments such as color-blindness or other conditions which might impair their ability to properly see the visual stimuli. Of the 65 participants who wore corrective lenses, three participants wore glasses with lenses that interfered with data capture with of eye-tracking equipment. Although we did not record fixations counts or fixation durations for these individuals, they were still able to provide valuable information and participated in all phases of the project. One examiner was unable to complete the entire data collection due to a schedule conflict. As with the Lay participants, all the data available for this examiner were included in our analyses.

Materials and Equipment. Standardized survey instrument. The survey instrument contained closed- and open-ended questions concerning FDE participant education, training, experience, and certification.

Eye tracking protocol. All eye tracking protocols were conducted with Tobii binocular eye tracker systems with 17" TFT, 1280 x 1024 pixel displays (Tobii Technology, Stockholm, Sweden), and Tobii Studio software (version 3.4.5). Dell model Precision 7710 laptop computers with Intel Core i7-6820HQ CPUs at 2.7GHz, 8GB RAM and 64-bit Windows 7 operating systems were used with all eye tracking units. The eye tracker computers were equipped with Osprey capture cards.

Open-ended interviews were recorded using Olympus WS-853 8GB didital voice recorders. Digital recordings were transcribed using Start Stop Universal Transcription software by HTM Engineering.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Signature stimuli. Genuine and simulated signatures were produced by approximately 100 writers on Wacom Intuos 3 digitizing tablets with Wacom Intuos 3 inking pens and Neuroscript Movalyser software. This enabled capture of handwriting speed and pressure data for each signature. We classified signature types based on the number of allographs present and legible within the signature (Mohammed, Found, & Rogers, 2008; Nguyen, Hammond, & Salyards, 2011). For our purposes, this classification scheme was used by our research team professional document examiner consultants to identify (1) text-based signatures in which each allograph of the name is clearly written; and (2) stylized signatures, in which one or fewer allographs are legible.

We further classified signatures according to their complexity using Found and colleagues' method of determining the complexity of signatures by evaluating the number of turning points, line intersections, and retrace strokes in a signature. The FDE expert consultants evaluated signature samples and selected text based and stylized signatures of high and low complexity. From these we created a final total of 56 signature comparisons incorporating high and low signature complexity, and text based and stylized signature types characteristic of the range of writings FDEs might encounter in casework. Signatures were scanned into a computer using Adobe Photoshop and saved as 8-bit grayscale 1024 x 768-pixel jpeg files (10 maximum quality) to enable their display on the Tobii binocular eye tracking systems.

Eye tracking Protocols. Participants were calibrated to a 9-point reference grid, which provided a resolution of subject gaze to better than 0.5 degree of visual angle. Participants viewed the eye tracker screen from 57 cm away so that the visual angle of the screen was 331 x 271 (W x H). The width of a typical questioned signature subtended a visual angle of approximately 281. Eye fixations were defined by the Tobii Studio software (version 3.4.5) as the eye position remaining within a 50-pixel area for a time of greater than 100 msec. Data collection, fixation measurements, and analysis of areas of interest (AOI) data were determined using Tobii Studio software. AOIs were defined prior to data analysis to analyze participant attention to different features within the signatures. All analyses to date have been performed at a macro (entire signature) level for the 56 individual signatures. Micro-level (features within signatures) analyses will be deferred for future publications.

To address the context question, we varied the presentation order of the questioned signature and the known signatures, using several formats, as described below. Signatures varied by type and complexity across protocols. In this experimental protocol the questioned signature was presented prior to the questioned/known comparison stimulus (Q→Q/K) or the known signatures preceded the questioned/known stimulus, (K→QK). A sample of the protocol stimuli is included in Appendix A. Participants viewed 5 sets of 4 signatures per set (20 comparisons).

In Experiment 2, four signatures were presented singly, sequentially (Q→Q→Q→Q), simultaneously (QQQQ), as demonstrated in Appendix B.

We defined sufficiency as both the number of writings available for comparison and the amount of complexity within the writings, as indicated by the number of turning points (5 TP; 10 TP; 20 TP, 30 TP, and 40 TP). In the sufficiency protocols (format 3) the questioned signature was paired with increasing numbers of known signatures (Q alone; Q+1K; Q+2K; Q+3K…Q+6K, as demonstrated in Appendix C). Comparisons were also presented in a similar format to those in Experiment 1.

After each comparison participants were asked to give an opinion about whether the questioned signature was genuine (written by the same person who wrote the known signatures) or simulated (written by different people). They were also asked to give their opinion about their Opinion Strength on the commonly used 9 level scale ranging from Inconclusive to Identification/Elimination. Finally, they were asked to give the certainty of their decision on a 20 point Likert-type scale ranging from 1 (not at all certain) to 20 (extremely certain). The scale levels were unmarked, and Decision Certainly was recorded by mouse click as the experimenter scrolled the mouse over the scale positions until the participant indicated that the cursor had reached their level of certainty.

Decision Making Analyses. Twelve questioned/known comparisons from experiment 1 were displayed a second time. Participants were digitally recorded as they described the algorithm they used to reach their conclusions for each of the signature comparisons. They were asked to discuss which features figured most prominently in their decisions, and why. As the participant explained their decision making process, the researcher marked the location of the features being described on a grid sheet which displayed the stimulus signatures on a grid sheet with X/Y coordinates to ensure accuracy when the recordings were transcribed and the data were coded. An example of a grid sheet and transcription are presented in Appendix D.

Content Analysis of Decision Making Analysis. Most of the qualitative data obtained from the semi-structured interview transcripts has been content analyzed using classic systematic quantitative and qualitative content analysis procedures described in Holsti (1969). The unit of analysis for the content analysis was the transcript for each participant. The unit of observation was the transcript for each separate signature comparison within the transcript. Thematic codes were empirically revised throughout the coding process. Coding guidelines (e.g., mutually exclusive and exhaustive categories) described in Holsti (1969) and other well-known research methodology texts were employed. The coding accommodated multiple mentions (e.g., more than one code per variable). When the coding is completed, frequencies will be calculated for all categorical variables, and codes will be collapsed into broader conceptual categories if thematically and theoretically appropriate.

Coding Procedure. Three coders per transcript were used to accommodate coding, check-coding, and check-code verification. Coders received training on the coding scheme and the methods and procedures of content analysis and were trained on the definitions of all variables. One hundred percent of the units of observation will be check-coded and check-verified.

Analyses and Results to Date. Analyses are ongoing, but our key project findings are reported below.

Analysis 1: Total Fixation Duration by Participant Type and Stimulus Presentation Method. A 2 (Participant Type) x 5 (Stimulus Presentation Method) factorial ANOVA revealed significant main effects for Participant Type, F (1, 5363) = 311.86, $p < .001$, partial $\eta^2 = .055$, and Stimulus Presentation Method, F (4, 5363) = 71.45, $p < .001$, partial $\eta^2 = .051$. A significant interaction effect was also revealed for Participant Type x Stimulus Presentation Method, F (4, 5363) = 17.84, $p < .001$, partial $\eta^2 = .013$ (see Figure 1).

Post hoc analysis using *Tukey's HSD* revealed that mean fixation count was significantly different in the Questioned First condition compared to the Known First, Sequential, Simultaneous, and Sequential Unmaksing conditions, $p < .001$. The Known First condition was significantly different from the Sequential condition and the Sequential Unmaksing condition, $p < .001$. The Simultaneous condition was also significantly different from the Sequential Unmasking condition, $p < .001$. No significant differences were observed between the Known First, Sequential, or Simultaneous conditions. Table 1 presents the means and standard deviations for this comparison.

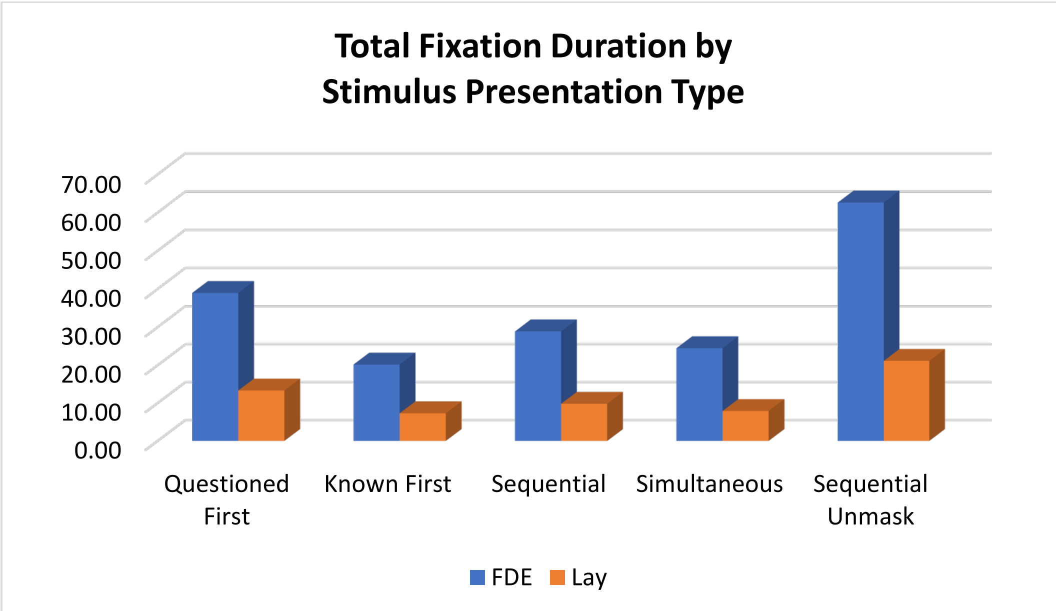


Figure 1. Total fixation duration (the total amount of time participants spent viewing the comparison stimuli) was longer for FDEs than for Lays at all levels of Stimulus Presentation Type. Both FDEs and Lays spent the most time viewing the Sequential Unmasking comparisons due to multiple presentations of pairings of the questioned and known signatures.

Table 1.
Fixation Duration by Participant Type and Stimulus Presentation Method

	FDE			Lay		
Stimulus Presentation	**N**	***M***	***SD***	**N**	***M***	***SD***
Questioned First*	849	38.94	42.67	244	13.33	17.39
Known First*	849	20.19	26.76	242	7.31	11.90
Sequential*	678	28.87	37.88	180	9.83	10.96
Simultaneous*	680	24.47	28.44	181	7.90	11.03
Sequential Unmask*	1182	62.63	57.15	288	21.15	30.17

* Significant at p < .001

Analysis 2: Total Fixation Count by Participant Type and Stimulus Presentation Method. A 2 (Participant Type) x 5 (Stimulus Presentation Method) factorial ANOVA revealed significant main effects for Participant Type, F (1, 5363) = 356.56, $p < .001$, partial $\eta^2 = .062$, and Stimulus Presentation Method, F (4, 5363) = 98.84, $p < .001$, partial $\eta^2 = .069$. A significant interaction effect was also revealed for Participant Type x Stimulus Presentation Method, F (4, 5363) = 24.43, $p < .001$, partial $\eta^2 = .018$ (see Figure 2).

Post hoc analysis using *Tukey's HSD* revealed that mean fixation count was significantly different in the Questioned First condition compared to the Known First, Sequential, Simultaneous, and Sequential Unmaksing conditions, $p < .001$. The Known First condition was significantly different from the

Sequential Unmaksing condition, $p < .001$. No significant differences were observed between the Known First, Sequential, or Simultaneous conditions. Table 2 presents the means and standard deviations for this comparison.

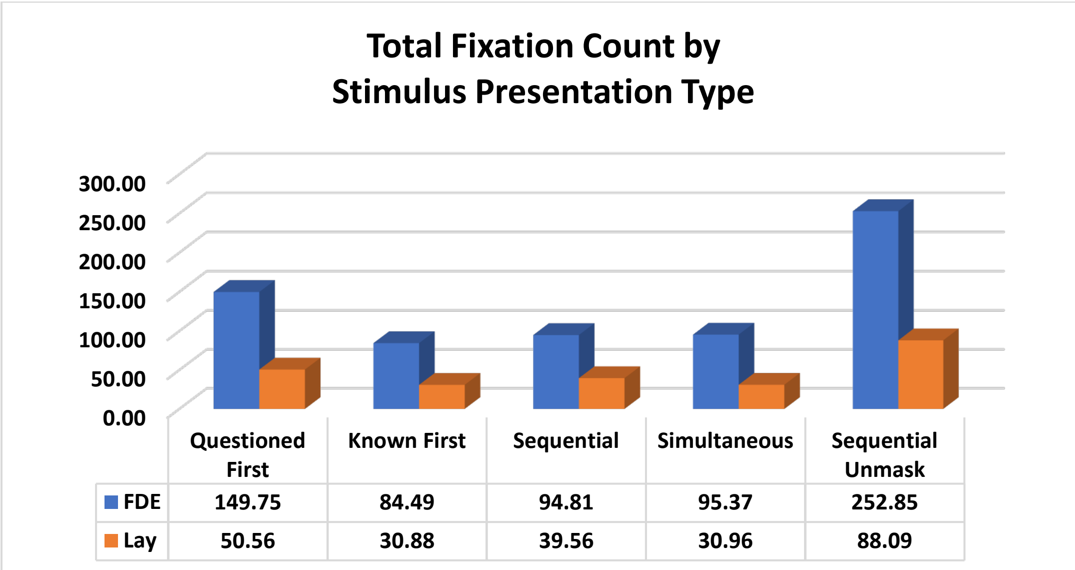


Figure 2. Total fixation count (the total number of times that the participant's eye position remained within a 50-pixel area for a time of greater than 100 msec) was greater for FDEs than for Lays at all levels of Stimulus Presentation Type. As seen in the fixation duration analysis, both FDEs and Lays recorded a greater number of fixations viewing the Sequential Unmasking comparisons due to multiple presentations of pairings of the questioned and known signatures.

Table 2.
Fixation Duration by Participant Type and Stimulus Presentation Method

	FDE			Lay		
Stimulus Presentation	**N**	***M***	***SD***	**N**	***M***	***SD***
Questioned First*	849	149.75	162.30	244	50.56	55.04
Known First*	849	84.49	103.61	242	30.88	36.94
Sequential*	678	94.81	89.40	180	39.56	39.39
Simultaneous*	680	95.37	104.13	181	30.96	40.75
Sequential Unmask*	1182	252.85	205.46	288	88.09	113.09

Analysis 3: Accuracy by Participant Type. We conducted a chi-square analysis to investigate the accuracy of FDE and Lay participants. FDEs made 3,725 correct decisions out of 4,747 total observations (78.5% accuracy, compared to 866 correct decisions out of 1,270 total observations made by Lay participants (68.2% accuracy), $\chi^2(1) = 58.57$, $p < .001$.

Analysis 4: Accuracy by Stimulus Presentation Method. Binomial logistic regression was conducted using the Enter method to determine whether Participant Type or Stimulus Presentation Method predicted Decision Accuracy (Accurate or Misleading). Regression results indicated that the overall model fit was questionable (-2 Log Likelihood = 5830.67), but was statistically reliable in distinguishing between decision types, $\chi^2(5) = 256.23$, $p < .001$. The model correctly classified 79.6% of cases. Figure 3 presents the percentage of accurate decisions by Stimulus Presentation Method.

Regression coefficients are presented in Table 3. The *Wald* statistic indicated that Participant Type significantly predicted whether the decision was accurate. However, the odds ratio for Participant Type was small, indicating little change in the likelihood of accuracy related to this factor. *Wald* statistics also indicated that Stimulus Presentation Method significantly predicted accuracy at all levels, although the odds ratios for the Known First and the Sequential presentation methods were also small, indicating little change in the likelihood of accuracy for those conditions.

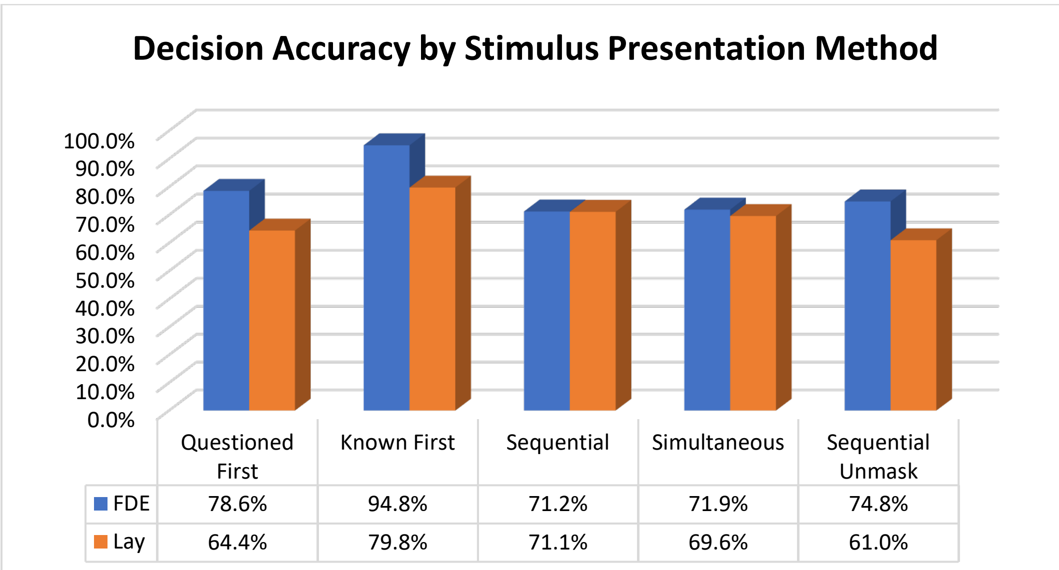


Figure 3. Stimulus Presentation Method and Participant Type were significantly related to Decision Accuracy. In all conditions FDEs were more accurate than were Lays, although we observed little difference in the percent of accurate calls in the Sequential and Simultaneous conditions.

Table 3.
Logistic Regression Coefficients for Participant Type and Stimulus Presentation Method

	B	*Wald*	*df*	*p*	*Odds*
Participant Type	-0.627	163.99	1	0.000	0.534
Stimulus Sequence		6.16	4	0.000	
Stimulus Sequence 1 (Known First)	-0.221	79.38	1	0.013	0.802
Stimulus Sequence 2 (Sequential)	-1.115	14.91	1	0.000	0.328
Stimulus Sequence 3 (Simultaneous)	0.378	14.05	1	0.000	1.459
Stimulus Sequence 4 (Sequential Unmask)	0.367	93.41	1	0.000	1.443
Constant	2.506	13.59	1	0.000	0.449

Analysis 5: Opinion Strength and Decision Certainty. We conducted a crosstabulation using the commonly used 9-level Opinion Strength continuum measure and Decision Certainty, a Likert-type scale ranging from 1 (not at all certain) to 20 (extremely certain). Results revealed

significant differences in between FDEs and Lays in the use of the Opinion Strength measure. The distribution of Opinion Strength on the Decision Certainty measure is presented in Figure 4.

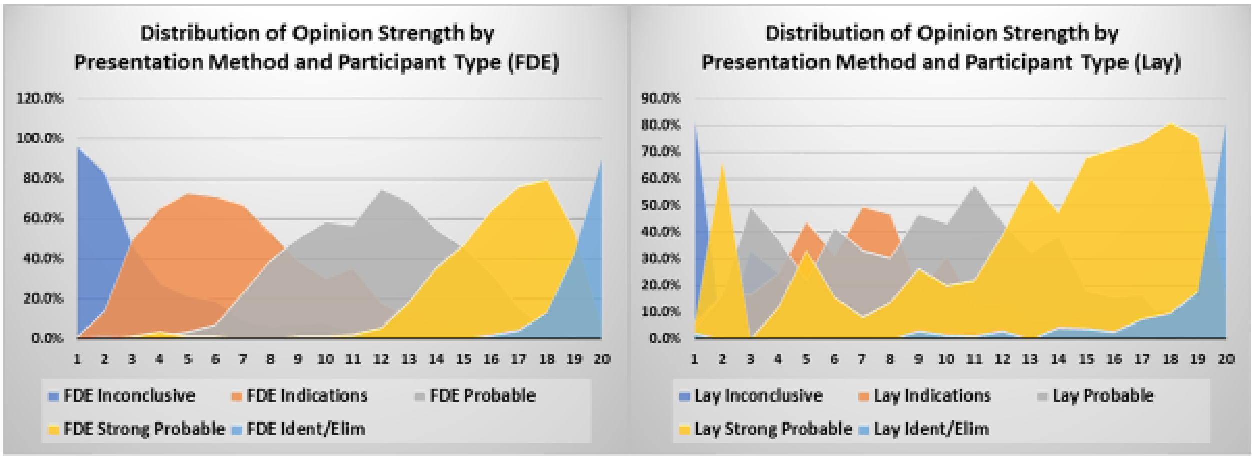
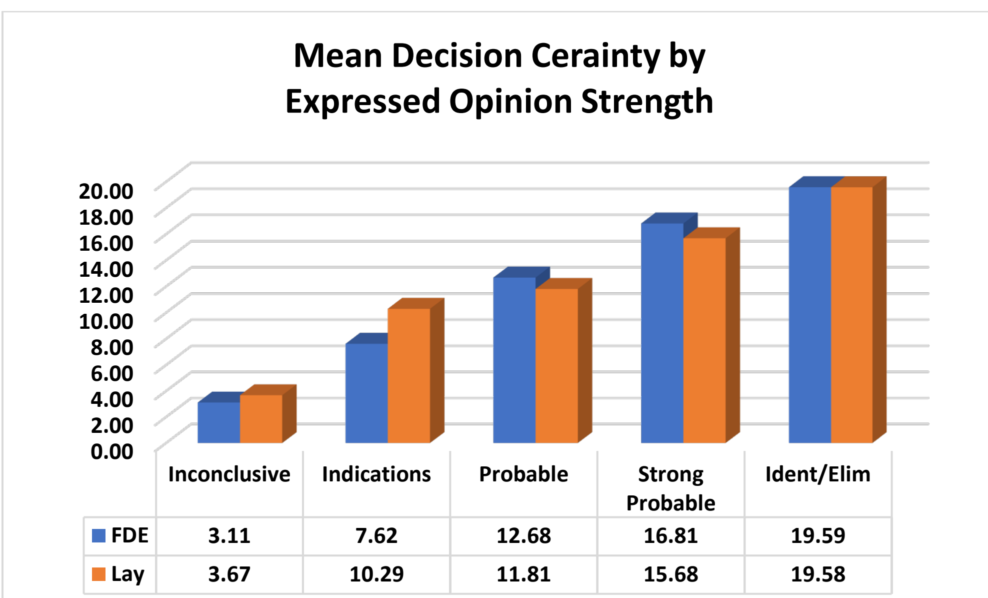


Figure 4. FDEs were more consistent than were Lays in their application of the Opinion Strength measure. This figure illustrates five clear levels of the Opinion Strength measure on the left, while on the right these levels are less evident.

We then conducted a 2 (Participant Type) x 5 (Opinion Strength) factorial ANOVA to investigate whether any significant differences existed in the measures of central tendency for the Opinion Strength continuum. We found significant main effects for Participant Type, F (1, 6007) = 4.69, p < .001, partial η^2 = .001, and Opinion Strength, F (4, 6007) = 2323.21, p < .001, partial η^2 = .607 (see Figure 4). We also found a significant interaction effect for Participant Type and Opinion Strength, F (4, 6007) = 38.90, p < .001, partial η^2 = .025 (see figure 5).



	Inconclusive	Indications	Probable	Strong Probable	Ident/Elim
FDE	3.11	7.62	12.68	16.81	19.59
Lay	3.67	10.29	11.81	15.68	19.58

Figure 5. This figure presents the mean Decision Certainty for comparisons across all protocols reported by FDEs and Lays for each level of Opinion Strength. This analysis reveals that except for the greater average Decision Certainty reported by Lays at the Indications level of the Opinion Strength measure, the centers of the distributions (means) demonstrate a systematic application of the Decision Certainty measure.

Table 4 presents data on the measures of central tendency. Although the centers of the distributions for FDEs and Lays are close, the dispersion of the data was less consistent for Lays than for FDEs, as demonstrated below in Figure 6.

Table 4.
Measures of Central Tendency by Opinion Strength for FDE and Lay Decision Certainty

FDE (n = 4,747)	**Inconclusive**	**Indications**	**Probable**	**Strong Probable**	**Ident/Elim**
N Observations	n = 613	n = 884	n = 1343	n = 1797	n = 1380
Mean	***3.17***	***7.96***	***12.52***	***16.50***	***19.58***
Median	1.00	7.00	13.00	17.00	20.00
Mode	1.00	5.00	12.00	18.00	20.00
Std Dev	3.44	3.76	3.13	2.79	1.38
Range	19.00	19.00	19.00	19.00	19.00
Minimum	1.00	1.00	1.00	1.00	1.00
Maximum	20.00	20.00	20.00	20.00	20.00
LAY (n = 1,270)	**Inconclusive**	**Indications**	**Probable**	**Strong Probable**	**Ident/Elim**
N Observations	n = 61	n = 113	n = 245	n = 495	n = 356
Mean	***3.67***	***10.29***	***11.81***	***15.68***	***19.58***
Median	1.00	10.00	11.00	16.00	20.00
Mode	1.00	10.00	11.00	17.00	20.00
Std Dev	4.37	4.03	3.62	3.67	1.69
Range	19.00	19.00	19.00	19.00	19.00
Minimum	1.00	1.00	1.00	1.00	1.00
Maximum	20.00	20.00	20.00	20.00	20.00

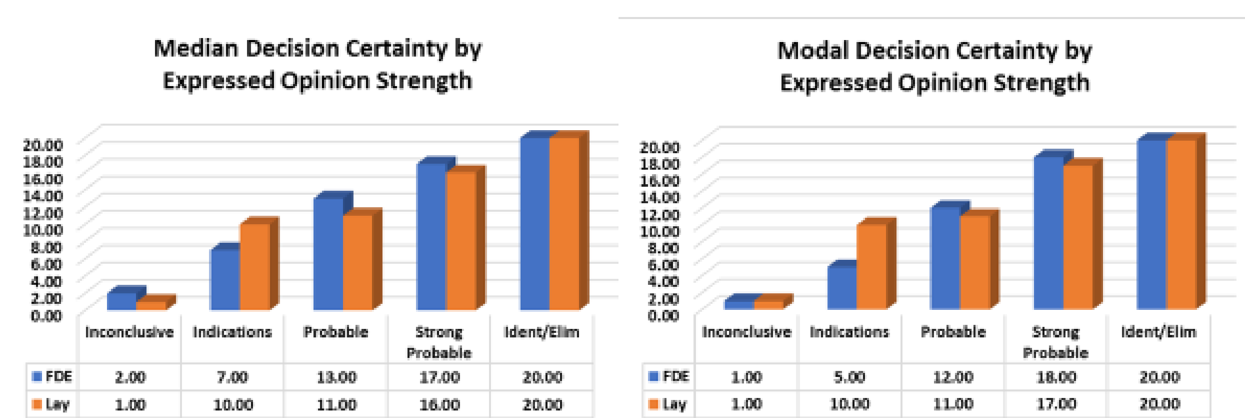


Figure 6. Standard deviations at all levels of the Opinion Strength measure were larger for Lay participants than for FDEs. This finding should be interpreted cautiously, however, given the relatively small sample size for the Lay group, which may impact the variability of these data.

Analysis 6: Opinion Strength and Decision Certainty and Signature Type. We conducted a 2 (Participant Type) x 2 (Signature Type) x 5 (Opinion Strength) factorial ANOVA to investigate whether the use of the Opinion Measure differed according to whether the signature was Text Based or Stylized. Significant main effects were found for Participant Type, Opinion Strength, and Signature Type, F (1, 5997) = 7.19, p = .007, partial η^2 = .001, F (4, 5997) = 1913.47, p < .001, partial η^2 = .561, F (1, 5997) = 4.13, p < .042, partial η^2 = .001, respectively. A significant interaction effect was found for Participant Type x Signature Type, F (4, 5997) = 39.94, p < .001, partial η^2 = .026. No significant interaction effects were found for Participant Type x Signature Type, and Opinion Strength x Signature Type. The three-way interaction effect for Participant Type x Opinion Strength x Signature Type was not significant (see figure 7).

Post hoc analyses using *Tukey's HSD* revealed that all pairwise comparisons for Opinion Strength were statistically significant at the p < .001 level. Table 5 presents means and standard deviations for these analyses.

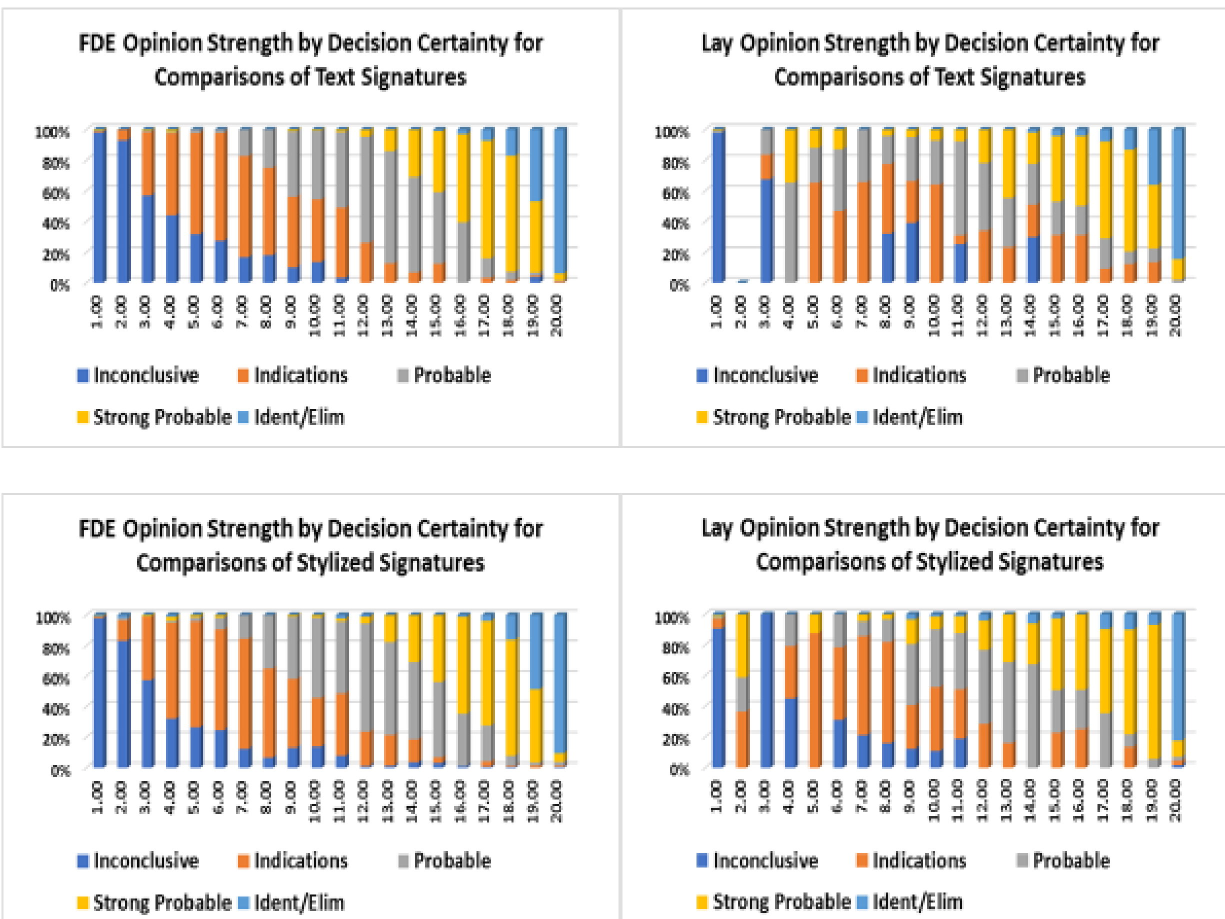


Figure 7. This figure illustrates the distribution of Decision Certainty over levels of Opinion Strength for Text Based and Stylized signatures. The charts indicate that the Opinion Strength measure was consistently applied by FDEs for both signature types, while Lay participants demonstrated more variability in its application.

Table 5.
Mean Decision Certainty and Opinion Strength for Text Based and Stylized Signatures

	Text Based Signatures					
	FDE			Lay		
	N	*M*	*SD*	N	*M*	*SD*
Inconclusive	149	3.12	3.06	11	4.64	4.90
Indications	333	7.72	3.61	47	11.04	3.75
Probable	543	12.68	2.84	136	11.60	3.63
Strong Probable	702	16.87	2.18	259	15.68	3.41
Ident/Elimination	563	19.63	0.77	175	19.71	0.95
	Stylized Signatures					
	FDE			**Lay**		
	N	*M*	*SD*	N	*M*	*SD*
Inconclusive	403	3.11	3.42	50	3.46	4.26
Indication	438	7.55	3.58	66	9.76	4.17
Probable	555	12.68	3.12	109	12.06	3.60
Strong	600	16.75	2.43	236	15.69	3.93
Ident/Elimination	461	19.53	1.68	181	19.46	2.18

Analysis 7: Opinion Strength, Decision Certainty and Identification/Elimination of Writers. We conducted a 2 (Participant Type) x 2 (Examination Opinion) x 5 (Opinion Strength) factorial ANOVA to investigate whether the use of the Opinion Measure differed according to whether the participant believed the signature to be Genuine or Simulated. Significant main effects were found for Participant Type, Opinion Strength, and Examination Opinion, F (1, 5997) = 9.02, p = .003, partial η^2 = .002, F (4, 5997) = 1999.26, p < .001, partial η^2 = .571,
F (1, 5997) = 18.11, p < .001, partial η^2 = .003, respectively.

A significant interaction effect was found for Participant Type x Opinion Strength, F (4, 5997) = 37.72, p < .001, partial η^2 = .025. A significant interaction effect was found for Participant Type x Examination Opinion, F (1, 5997) = 7.45, p = .006, partial η^2 = .001. A significant interaction effect was also found for Opinion Strength x Examination Opinion, F (4, 5997) = 3.63, p = .006, partial η^2 = .002. The three-way interaction effect for Participant Type x Opinion Strength x Examination opinion was statistically significant, F (4, 5997) = 3.82, p = .004, partial η^2 = .003 (see figure 8).

Post hoc analyses using *Tukey's HSD* revealed that all pairwise comparisons for Opinion Strength were statistically significant at the p < .001 level. Table 5 presents means and standard deviations for these analyses.

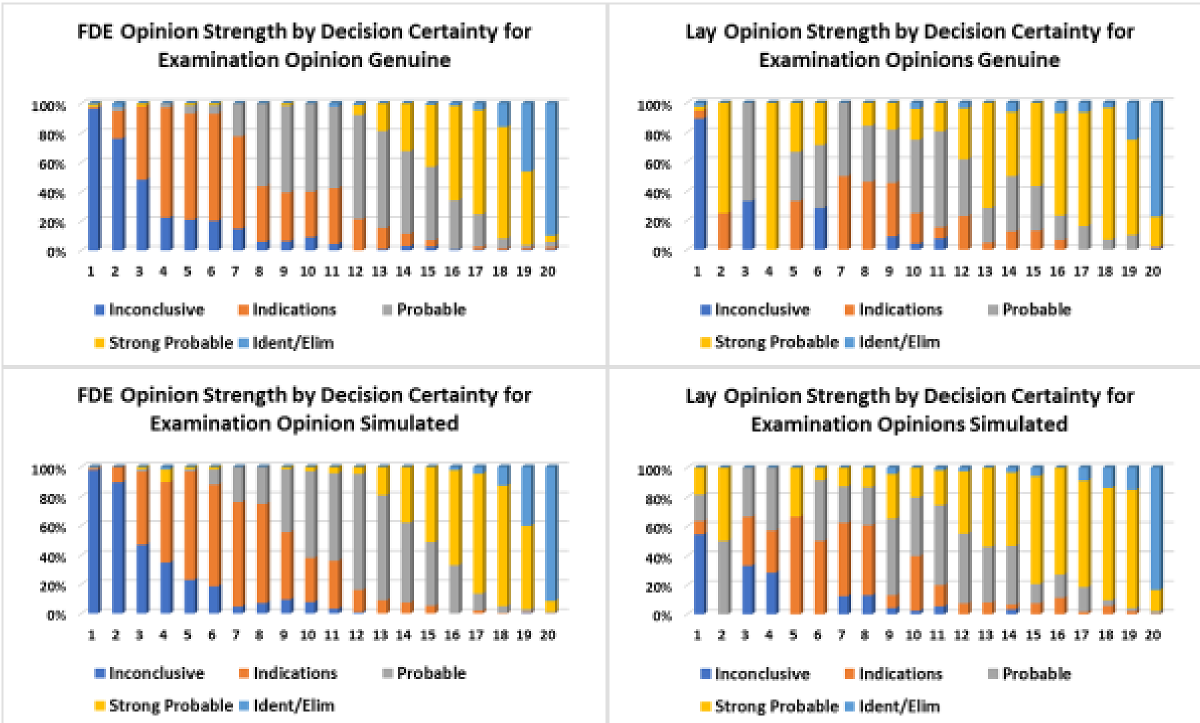
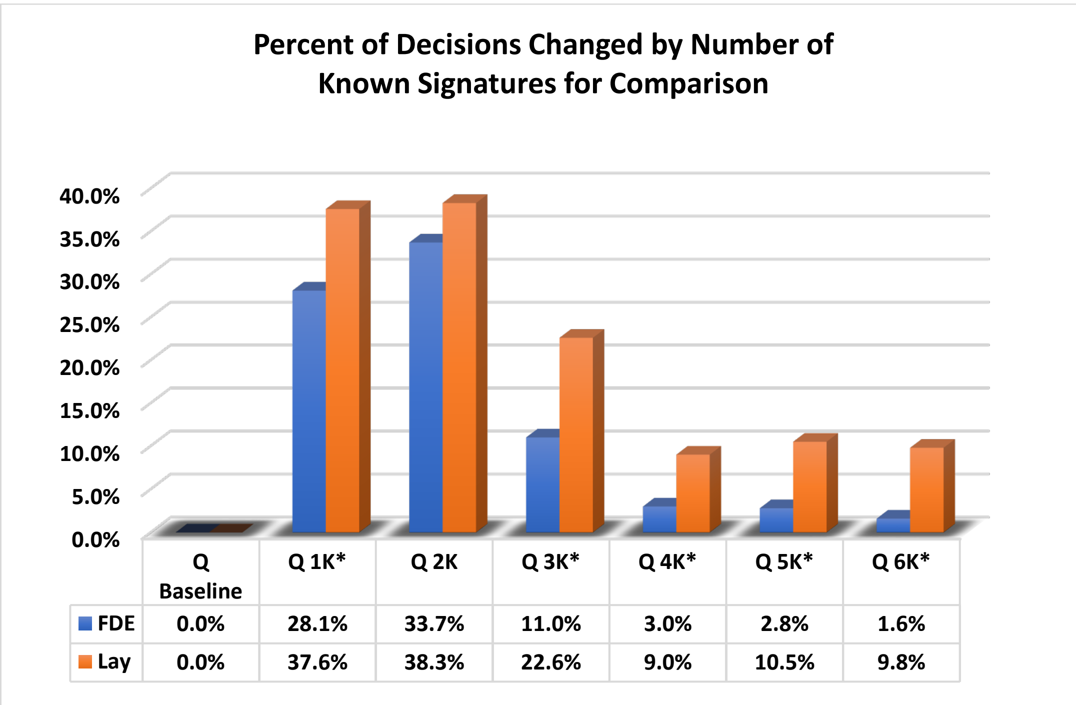


Figure 8. This figure illustrates the distribution of Decision Certainty over levels of Opinion Strength for Examination Opinions of genuine or simulated. Similar to the findings for Signature Type, the charts indicate that the Opinion Strength measure was more consistently applied by FDEs for both signature types, while Lay participants demonstrated more variability in its application.

Table 5.
Mean Decision Certainty and Opinion Strength for Genuine and Simulated Examination Opinions

	Genuine Opinion					
	FDE			**Lay**		
	N	*M*	*SD*	N	*M*	*SD*
Inconclusive	324	3.06	3.62	42	2.62	3.98
Indications	387	3.74	3.61	39	10.10	4.09
Probable	525	12.63	3.16	90	11.82	3.56
Strong Probable	523	16.73	2.31	176	15.49	4.04
Ident/Elimination	346	19.35	1.76	110	19.42	2.31
	Simulated Opinion					
	FDE			**Lay**		
	N	*M*	*SD*	N	*M*	*SD*
Inconclusive	228	3.18	2.86	19	6.00	4.37
Indication	384	7.72	3.44	74	10.39	4.03
Probable	573	12.72	2.81	155	11.80	3.66
Strong	779	16.87	2.29	319	15.79	3.45
Ident/Elimination	678	19.71	0.89	246	19.65	1.32

Analysis 8. Relationships Between Signature Complexity and Quantity and Decision. We conducted a correlation analysis using the *Person's Product Moment* statistic to investigate the relationship between Signature Complexity as measured by the number of turning points (e.g., the number of times the pen changed direction) in the writing and the number of Examination Decision Changes as participants viewed a series of comparisons of the same questioned signature. The number of changes was significantly positively correlated to the number of turning points, r (640) = .143, p < .001. Figure 9 demonstrates the percent of decisions changed as the number of known signatures available for comparison increased. Figure 8 presents



	Q Baseline	Q 1K*	Q 2K	Q 3K*	Q 4K*	Q 5K*	Q 6K*
FDE	0.0%	28.1%	33.7%	11.0%	3.0%	2.8%	1.6%
Lay	0.0%	37.6%	38.3%	22.6%	9.0%	10.5%	9.8%

Figure 9. The greatest amount of Examination Decision change occurred after the baseline condition in which the questioned signature was displayed without any known signatures for comparison. The number of known signatures displayed increased with each comparison. Little decision change was observed after participants had viewed the questioned signature paired with 3 known signatures (Q 3K). A greater percentage of Lay participants than of FDES changed decisions across all levels of known signatures available for comparison.

We conducted a series of independent group t tests to investigate whether Opinion Strength differed by Decision Change (change/no change) after each comparison. A significant difference was found for FDEs only in the Q 3K comparison, for which the mean opinion strength was higher for no change (m = 2.33) than for change (m = 2.11), t (240.21) = 1.98, p = .048. Mean opinion strength for Lays was significantly higher in the Q 4K comparison for no opinion change (m = 3.98) than for change (m = 3.62), t (130) = 2.10, p = .038. Figure 10 demonstrates the distribution of Examination Decision change by Opinion Strength and number of known signatures for comparison.

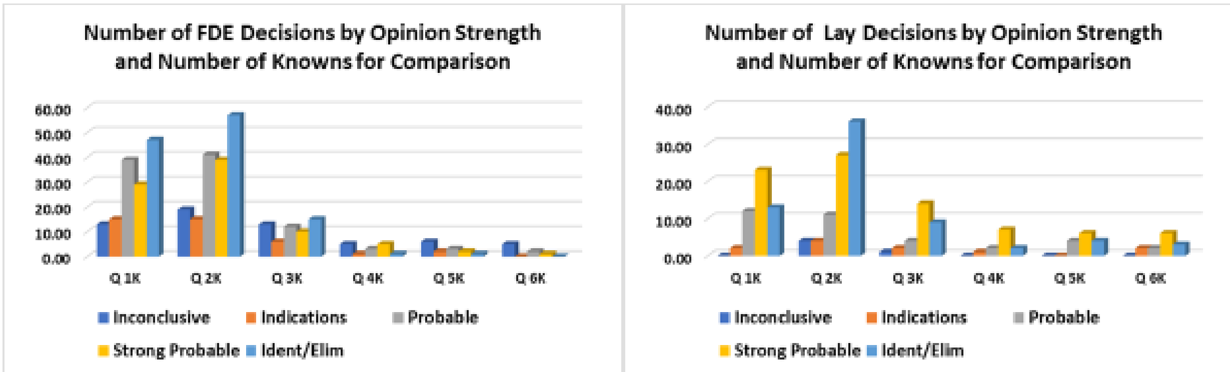


Figure 10. This figure compares the distribution of Examination Opinion changes for each pairing of questioned and known signatures. After the third comparison (Q 3K) there is little decision change among FDEs compared to Lay participants.

Discussion

Taking the findings above in the context of our research questions and hypotheses, we offer the following conclusions.

Research Question 1 (Experiment 1): How is the presentation of questioned and known signatures during a signature comparison related to the determination of signature and opinion decision outcomes?

Hypothesis 1a: Compared to simultaneous presentation of the signatures, sequential presentation of the signatures will result in a more extensive examination of each signature, as indicated by the total number of reference saccades per signature. ***This hypothesis was not supported.*** We did find differences in fixation count and fixation duration in Analysis 1 and Analysis 2 but the specific hypothesis about sequential vs. simultaneous presentation of stimuli was not supported.

Hypothesis 1b: Sequential presentation will result in greater accuracy for genuine signatures than for simulated signatures. ***This hypothesis was supported.***

Hypothesis 1c: Simultaneous presentation will result in greater accuracy than will sequential presentation. ***This hypothesis was supported.*** Analysis 4 demonstrated that Stimulus Presentation Method was a significant predictor of Decision Accuracy, and that the Simultaneous presentation was related to greater accuracy, while greater inaccuracy occurred in the Sequential presentation.

Hypothesis 1d: Sequential presentation will result in less extreme confidence levels on the 9-position opinion continuum than will simultaneous presentation. ***This hypothesis was supported.*** Evidence for this was found by examining the measures of central tendency reported in Analysis 5.

Research Question 2 (Experiment 2): How do examiners apply the currently used bipolar continuum of certainty (Elimination through Identification with a center position of Inconclusive) when expressing their opinions about the authorship of questioned writings?

Hypothesis 2a: The intervals will vary such that the boundaries for intervals indicating low support for conclusions (inconclusive or indications) will be more widely dispersed than the intervals indicating higher levels of support (strong probable or identification/elimination), as measured by the values on the decision certainty scale. ***This hypothesis was supported.*** Evidence for this is found in the measures of central tendency reported in Analysis 5.

Hypothesis 2b: Use of the intervals on the 9-level certainty continuum will vary according signature type and complexity. ***This hypothesis was supported.*** Analysis 6 demonstrated support for this hypothesis this directly for signature type and complexity. Analysis 7 demonstrated the effect for use of the Opinion Strength measure for examination decisions of genuine (identification) or simulated (elimination).

Hypothesis 2bii: Dispersion of the interval boundaries will be greater for stylized than for text-based signatures. ***This hypothesis was supported.*** Analysis 6 revealed that FDEs use the Opinion Strength measure consistently for both signature types, while Lays appeared to be much more inconsistent.

Hypothesis 2c: These differences will be more pronounced for genuine (identification) than for simulated (elimination) signatures. ***This hypothesis was supported.*** Analysis 7 demonstrates that the results for this hypothesis are very similar to those in Analysis 6, in that FDEs used the Opinion Strength measure more consistently than did Lay participants.

Hypothesis 2d: Use of the continuum boundaries will differ between FDEs and Lays participants. ***This hypothesis was supported*** *(see above).*

Hypothesis 2e: FDE opinions will be more conservative than will Lay opinions. ***This hypothesis was supported*** *(see above).* The measures of central tendency in Analysis 5 demonstrate that FDEs make clear distinctions across levels of Opinion Strength, while Lays are particularly inconsistent at the first three levels (Inconclusive, Indications, and Probable).

Research Question 3 (Experiment 3): How much writing constitutes "sufficient" information upon which to base an opinion?

Hypothesis 3a: FDEs will be more conservative than will lay participants, such that they will require a greater amount of writing before making a process call of genuine or simulated. ***This hypothesis was supported.*** Analysis 8 demonstrated that after giving a baseline decision following examination of the questioned signature alone, 28% of FDE decisions were changed from the original decision the questioned signature was displayed with one known signature, compared to 37.6% of Lay decisions. This pattern persists across all comparison conditions.

What is the impact on other disciplines?

Dissemination to date and works in progress are presented in Appendix E. Practitioners from many forensic fields have taken seriously the need for standardized training and proficiency testing, and through organizations such as the NIST OSAC are working nationally and internationally to define and establish valid and reliable measures of certainty, proficiency, and error. Forensic experts around the world are striving to ensure that their methods are transparent to the courts, and that judges are given the information they need to make their decisions. Efforts to organize and present information effectively, which are important goals of OSAC, have been an important consequence of the *Daubert* trilogy and the NAS report. Forensic scientists are also seeking opportunities to collaborate with judges, attorneys, and scientists from other fields on research and education projects.

Preliminary findings from this research have been disseminated at several international venues (the American Academy of Forensic Sciences, the American Society of Questioned Document Examiners, and the European Network of Forensic Handwriting Experts). Our interactions with international experts have to potential to profoundly affect the scope and impact of our research by facilitating the creation of new professional relationships across the globe. By engaging examiners from Canada, Europe, and Australia, we have established new collaborations that will take our research to the international level. We have already submitted two additional proposals with international collaborators from Canada, Scotland, and Australia, and will continue to maintain our relationship with the Grant Sawyer Center for Justice Studies at the University of Nevada, Reno, where we will engage members of the judiciary attending classes at The National Judicial College pending funding of one of our projects currently under review.

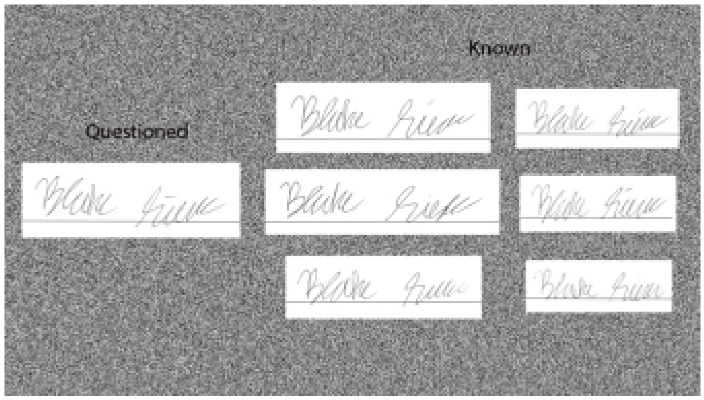
Our findings have implications not only for Questioned Examination, but also for other areas of Pattern and Physics Evidence identified by NIST OSAC and other organizations. Our research methods can be adapted to other disciplines, which will increase the understanding of cognitive human factors in those fields and provide information about possible sources of cognitive bias, such as the semantic context of the specimen, order of signature presentation, top-down/bottom-up processing of information, order of presentation effects, word-superiority effects, and other relevant cognitive phenomena. The movement of expert testimony from the status of “proffer” to that of “admissible evidence” is a social process in which experts, attorneys, and judges all participate. It is a negotiated movement from “science,” which is itself a social construction,[1] to “legal science,”[2] which is mediated by the rhetoric and discourse of attorneys, judges, and academicians. Transparency of methods is an important component of the admissibility of FDE testimony. Eye tracking methodology, physiological data, the diagnostic value of the evidential features of handwriting, and descriptions of the decision making process will help increase the transparency of the examination process, improving the quality of performance of attorneys, judges, and experts.

[1] Jasanoff, S. (1993). What judges should know about the sociology of science. *Judicature, 7*, 77-82.

[2] Caudill, D.S. (2001). Law and science: An essay on links and socio-natural hybrids. *Syracuse Law Review, 51,* 841-862.

Appendix A: Sample Experiment 1 Stimulus

Experiment 1: Questioned/Known Signature Position Comparison Trials



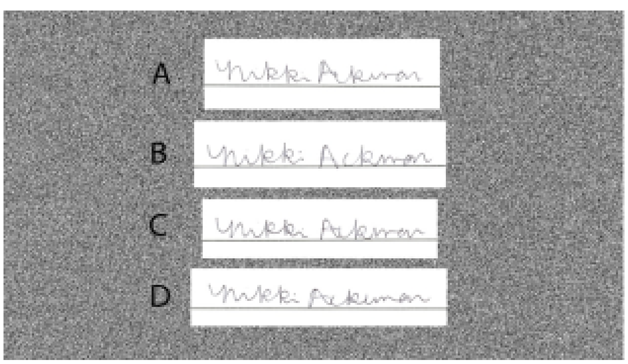
- Five sets, each consisting of four signatures (20 total comparisons)
- Process decision: Is the questioned signature is **genuine** (the signature is the true signature of the person who wrote it) or **simulated** (the writer of the questioned signature attempted to either copy or imitate the signature of a different person)?

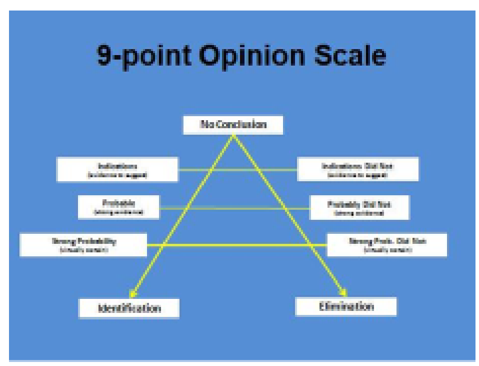
Appendix B: Sample Experiment 2 Stimuli

Experiment 2: Sequential Comparison Trials

- Four sets consisting of two signatures comparisons per set
- Did one writer, or more than one writer produce the signatures in the set
- Confidence in your decision based on the confidence scale
- How likely you think it is that the person who wrote the questioned signature is the same person who wrote the known signatures

Experiment 2: Simultaneous Comparison Trials

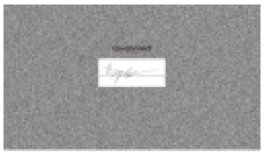


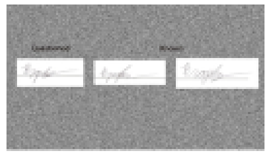


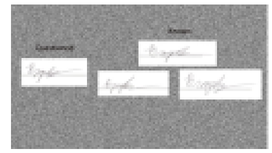
- Two sets consisting of four signatures comparisons per set
- Did one writer, or more than one writer produce the signatures in the set
- Confidence in your decision based on the confidence scale
- How likely you think it is that the person who wrote the questioned signature is the same person who wrote the known signatures

Appendix C: Sample Experiment 3 Stimuli

Experiment 3: Sufficient Writing Quantity and Complexity

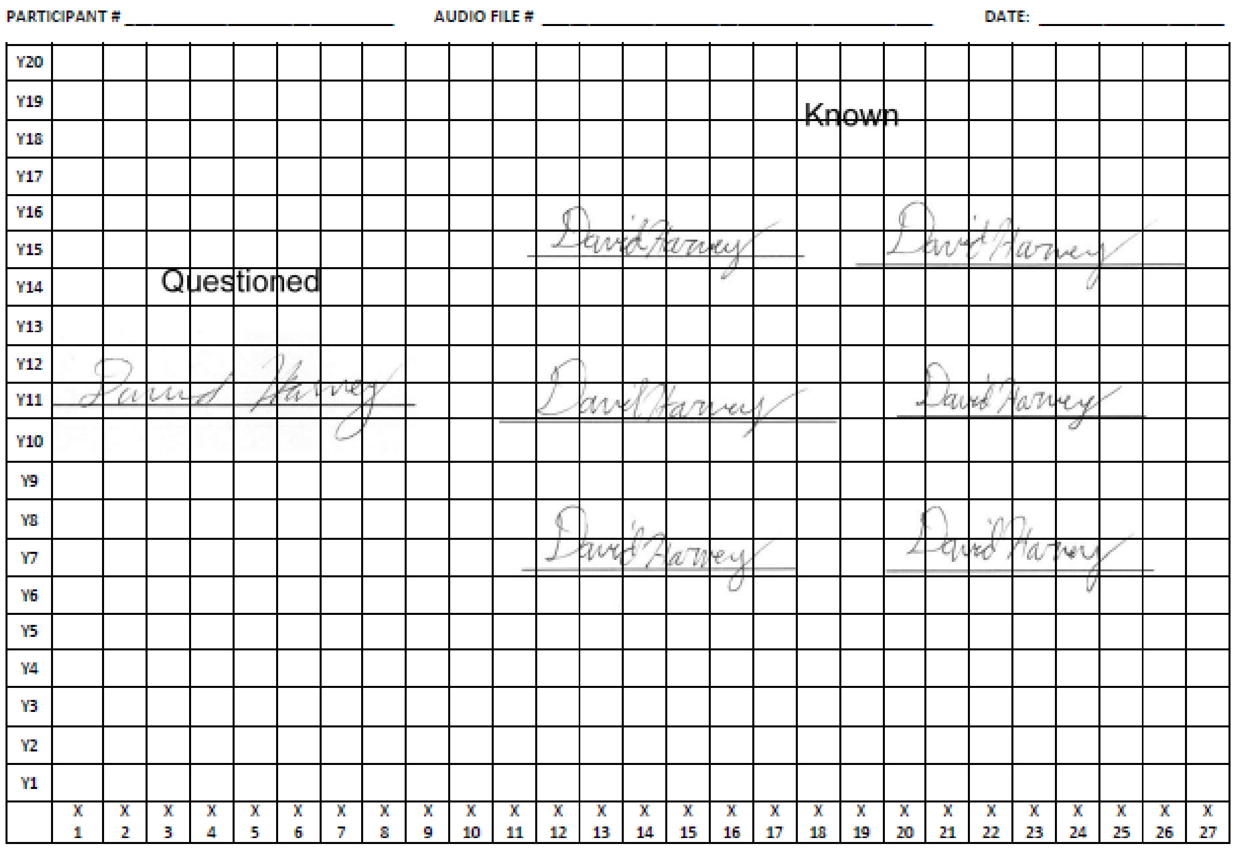






- Ten sets with 2 signatures per set
- The Questioned signature will be presented with a varying number of Known signatures
- The number of Known signatures increases with each view of the Questioned signatures
- View the Questioned signature first, then respond to the questions per the previous experiments
- Is the questioned signature genuine or simulated
- Confidence in your decision based on the confidence scale
- How likely you think it is that the person who wrote the questioned signature is the same person who wrote the known signatures
- Repeat this process for up to six Questioned/Known signature comparisons

Appendix D: Sample Signature Grid Sheet and Qualitative Interview Transcription



Transcription 20241 Signature 10 Harvey 161220-0084 7m 30s

Speaker 1: Mariah Bailey

Speaker 2: 20241

Transcriber: Merlino

Harvey

Speaker 2: There were a couple of characters that stood out to me with some characteristics that were dissimilar between the questioned and the known. The first is the capital letter D in the questioned document it has a very large loop at the bottom of the D seen here

Speaker 1: Circled y12/x1

Speaker 2: And it also has a smaller body of the D, which you can see here

Speaker 1: Circled y13/x2

Speaker 2: Now when compared to the known signatures, a couple of them did have loops at the base of the D but they were very tiny, for example here

Speaker 1: Circled y12/x20

Speaker 2: And then the body of the D is the majority of the character itself, seen here

Speaker 1: Circled y16-27/x20

Speaker 2: So again like the knowns, the body of the D is the majority part compared to the question where it's almost like a figure 8, where the bottom loop is quite large and the body of the D is almost similar in size. The other thing that I paid attention to was the connecting stroke from the crossbar of the H in the last name to the lower case A which can be seen in the questioned

Speaker 1: Circled y12/x6

Speaker 2: Where in all of the knowns the crossbar of the H is separate from the top of the letter A, so the crossbar of the H seen here

Speaker 1: Circled y8/x14

Speaker 2: Does not connect the top of the A seen there

Speaker 1: Y8/x15

Speaker 2: Another characteristic that stuck out to me was the letter R in the last name. In the questioned document it is a peaked apex, it almost has a tiny retrace, a little tick mark at the top seen here in the questioned

Speaker 1: Circled y12-23/x6

Speaker 2: And then when comparing that to the known signatures the R is very box shaped with a flat top which can be seen here

Speaker 1: Compared to y12/x23

Speaker 2: Something else that I didn't put as much weight on but I did note was the descender of the letter Y in the last name. In the questioned document it's quite large in comparison to the bowl

Speaker 1: Circled y11-12/x7

Speaker 2: Where in the known signature it's pretty similar in size or standard in size with the bowl of the Y seen here

Speaker 1: Compared to y11-12/x24. Transition over to your heat map, there are some specific warm areas here. Can we talk a little bit about your focus and why you focused in those areas?

Speaker 2: I think the two biggest things that I focused on are the biggest, in the questioned was the connecting stroke from the H crossbar to the letter A, it's really difficult to do something like that or to not do something like that if it's not natural to you, and that I didn't notice that in any of the known signatures, I looked at each one and I did not see any instance where the crossbar where the crossbar connected to the letter A. I think that was my biggest focus in the questioned signature, and then in looking at the knowns I did make sure to look at all of the Rs, as I had mentioned before, which are a little box shaped compared to the questioned document.

Speaker 1: Here in the questioned signature you have a focus a little bit here in the spacing, and I kind of noticed that same consistency in looking at the other signatures, so was that something that was a determining factor as well, or is that something that you kind of glanced but it didn't weigh heavy in your decision?

Speaker 2: No, I actually was wondering what I was looking at there. When I was looking at this heat map I wasn't sure what that was, but yes, that is something looking at this hard copy now and that previously looking at the blank signatures there is definitely a difference in spacing between the first and last name, so you can see like here in the questioned

Speaker 1: Circled y12-13/x5

Speaker 2: And that's something that I would have taken into consideration and observed as a dissimilarity, so the known signatures are almost all on the verge of touching, which you can see here

Speaker 1: Circled y8-9/x22

Speaker 2: That's it, I think.

Speaker 1: Anything else specific? We pretty much hit on all of the other hot areas on this map, but is there anything else in specific than being able to look at your heat map and see where your focus was, that you would want to talk about or hit on?

Speaker 2: No, there's not much on the heat map regarding the letter D, but it was quickly observed, you could see it easily. Other than that I think that the areas where there is a lot of focus is what helped me make my decision the most.

Speaker 1: Anything else? That concludes signature 10. [END TRANSCRIPT] 18m 30s

Appendix E: Dissemination to Date and Works in Progress

Merlino, M.L., (edited book chapter to be included in the Scientific Examination of Questioned Documents (2 Ed.), Taylor Francis (CRC) Publishers). *Cognitive Human Factors and Bias in Handwriting Examination: Concepts, Current Knowledge, and Future Research Directions.* [Submitted]

Merlino, M.L., Dahir, V.B., Hammond, D.L., Smith, L.R., Al Namer, C.D., Al Namer, T.A., Alvarez, M., Villalobos, J.G., & Edwards, C. (in preparation). *Cognitive human factors and forensic document examiner methods and procedures: Key results from an international study of handwriting experts.*

Merlino, M.L., Al Namer, C.D., Hammond, D.L., Olson, R.S., & Easley, P. (in preparation). *A comparison of gaze behavior in sequential versus simultaneous presentation of signatures in a handwriting comparison task.*

Merlino, M.L., Dahir, V.B., Edwards, C., Alvarez, M., Sanchez, S., Villalobos, J.G., Swinger, C., Schaar-Buis, D. (in preparation). *Education and training in forensic document examination: Issues and ideas.*

Published Conference Proceedings

Merlino, M.L., Dahir, V.B., Hammond, D.L., Smith, L.R., Al Namer, C.D., Al Namer, T.A., Alvarez, M., Villalobos, J.G., & Edwards, C. (2019, February). *Cognitive human factors and forensic document examiner methods and procedures: Key results from an international study of handwriting experts.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2019Proceedings.pdf

Merlino, M.L., Al Namer, C.D., Hammond, D.L., Olson, R.S., & Easley, P. (2019, February). *A comparison of gaze behavior in sequential versus simultaneous presentation of signatures in a handwriting comparison task.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2019Proceedings.pdf

Merlino, M.L., Alvarez, M., Al Namer, T.A., Villalobos, J.G, & Edwards, C. (2019, February). *Writing speed and fluidity and accuracy of calls in high and low complexity signature comparisons.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2019Proceedings.pdf

Merlino, M.L., Dahir, V.B., Edwards, C., Alvarez, M., Sanchez, S., Villalobos, J.G., Swinger, C., Schaar-Buis, D. (2019, February). *Education and training in forensic document examination: A discussion of issues and ideas.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2019Proceedings.pdf

Merlino, M.L. et al. (2017, February). *Interdisciplinary symposium: The evolution of Daubert and its effects on the forensic sciences.* Paper presented at the 2017 Annual Meeting of the American Academy of Forensic Sciences, New Orleans, LA. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2017Proceedings.pdf

Smith, L.R., & **Merlino, M.L.** (2017, February). *Questioned document examiner training: Foundational readings and resources.* Paper presented at the 2017 Annual Meeting of the American Academy of Forensic Sciences, New Orleans, LA. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2017Proceedings.pdf

Smith, L.R., & **Merlino, M.L.** (2017, February). *The admissibility of forensic expert testimony: 25 years of milestones and impacts.* Paper presented at the 2017 Annual Meeting of the American Academy of Forensic Sciences, New Orleans, LA. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2017Proceedings.pdf

Merlino, M.L., Freeman, T.M., Smith, L R., & Duvall, I. (2017, February). The background, training, and experience of questioned document examiners: Present practice and future directions. Paper presented at the 2017 Annual Meeting of the American Academy of Forensic Sciences, New Orleans, LA. Abstract retrieved from https://www.aafs.org/wp-content/uploads/2017Proceedings.pdf

Merlino, M.L., Freeman, T.M., Dahir, V., Springer, V., Hammond, D., Dyer, A., & Found, B.J. (2016, July). *Cognitive theoretical perspectives in studies of forensic document examination.* National Academies of Science Workshop on Personnel Selection tools for Forensic Sciences, Washington, DC. Abstract retrieved from https://www.nap.edu/catalog/23681/personnel-selection-in-the-pattern-evidence-domain-of-forensic-science

Invited Presentations

Merlino, M.L., Dahir, V., Hammond, D.L., Al Namer, C.D., Al Namer, T.A., Edwards, C.P., Schaar-Buis, D., Alvarez, M., Villalobos, J.G., & Smith, L.R. (October 2019). *Cognitive Human Factors and Forensic Document Examiner Methods and Procedures: Key Results from an International Study of Handwriting Experts.* Presented at the 2019 Annual Meeting of the Midwestern Academy of Forensic Sciences, Louisville, KY.

Merlino, M.L., Alvarez, M., Al Namer, T.A., Edwards, C.P., Hammond, D.L., & Dahir, V. (October 2019). *Writing Speed and Fluidity and Accuracy of Calls in High and Low Complexity Signature Comparisons.* Presented at the 2019 Annual Meeting of the Midwestern Academy of Forensic Sciences, Louisville, KY.

Merlino, M.L., Hammond, D.L., Al Namer, C.D., Al Namer, T.A., Edwards, C.P., Dahir, V. (August 2019). *Feature Salience and Call Accuracy in Simultaneous and Sequential Writing Comparison Tasks.* Presented at the 2019 Annual Meeting of the Midwestern Academy of Forensic Sciences, Louisville, KY.

Edwards, C.P., **Merlino, M.L.,** Dahir, V.B., Alvarez, M., Sanchez, C., Villalobos, J.G., Swinger, C., Schaar-Buis, D., Caufield, K.E. (October 2019). *Education and Training in Forensic Document Examination: A Discussion of Issues and Ideas.* [Phase I survey results.] Presented at the 2019 Annual Meeting of the Midwestern Academy of Forensic Sciences, Louisville, KY.

Merlino, M.L., Edwards, C.P., Dahir, V., Hammond, D.L., Alvarez, M., Sanchez, C., Villalobos, J.G., Swinger, C., Schaar-Buis, D., & Caufield, K.E. (October 2019). *Training in Forensic Document Examination: Current Thoughts and Future Directions.* [Phase III results.] Presented at the 2019 Annual Meeting of the Midwestern Academy of Forensic Sciences, Louisville, KY.

Merlino, M.L., Dahir, V., Hammond, D.L., Al Namer, C.D., Al Namer, T.A., Edwards, C.P., Schaar-Buis, D., Alvarez, M., Villalobos, J.G., & Smith, L.R. (September 2019). *Cognitive Human Factors and Forensic Document Examiner Methods and Procedures: Key Results from an International Study of Handwriting Experts.* Paper presented at the 2019 Biennial Meeting of the European Network of Forensic Handwriting Examiners, Porto, Portugal.

Merlino, M.L., Dahir, V., Hammond, D.L., Al Namer, C.D., Al Namer, T.A., Edwards, C.P., Schaar-Buis, D., Alvarez, M., Villalobos, J.G., & Smith, L.R. (August 2019). *Cognitive Human Factors and Forensic Document Examiner Methods and Procedures: Key Results from an International Study of Handwriting Experts.* Paper presented at the 2019 Annual Meeting of the American Society of Questioned Document Examiners, Cary, NC.

Merlino, M.L., Alvarez, M., Al Namer, T.A., Edwards, C.P., Hammond, D.L., & Dahir, V. (August 2019). *Writing Speed and Fluidity and Accuracy of Calls in High and Low Complexity Signature Comparisons.* Presented at the 2019 Annual Meeting of the American Society of Questioned Document Examiners, Cary, NC.

Merlino, M.L., Hammond, D.L., Al Namer, C.D., Al Namer, T.A., Edwards, C.P., Dahir, V. (August 2019). *Feature Salience and Call Accuracy in Simultaneous and Sequential Writing Comparison Tasks.* Presented at the 2019 Annual Meeting of the American Society of Questioned Document Examiners, Cary, NC.

Edwards, C.P., **Merlino, M.L.,** Dahir, V.B., Alvarez, M., Sanchez, C., Villalobos, J.G., Swinger, C., Schaar-Buis, D., Caufield, K.E. (2019). *Education and Training in Forensic Document Examination: A Discussion of Issues and Ideas.* [Phase I survey results.] Presented at the 2019 Annual Meeting of the American Society of Questioned Document Examiners, Cary, NC.

Merlino, M.L., Edwards, C.P., Dahir, V., Hammond, D.L., Alvarez, M., Sanchez, C., Villalobos, J.G., Swinger, C., Schaar-Buis, D., & Caufield, K.E. (August 2019). *Training in Forensic Document Examination: Current Thoughts and Future Directions.* [Phase III results.] Presented at the 2019 Annual Meeting of the American Society of Questioned Document Examiners, Cary, NC.

Merlino, M.L., Dahir, V.B., Hammond, D.L., Smith, L.R., Al Namer, C.D., Al Namer, T.A., Alvarez, M., Villalobos, J.G., & Edwards, C. (2019, February). *Cognitive human factors and forensic document examiner methods and procedures: Key results from an international study of handwriting experts.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD.

Merlino, M.L., Al Namer, C.D., Hammond, D.L., Olson, R.S., & Easley, P. (2019, February). *A comparison of gaze behavior in sequential versus simultaneous presentation of signatures in a handwriting comparison task.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD.

Merlino, M.L., Alvarez, M., Al Namer, T.A., Villalobos, J.G, & Edwards, C. (2019, February). *Writing speed and fluidity and accuracy of calls in high and low complexity signature comparisons.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD.

Merlino, M.L., Dahir, V.B., Edwards, C., Alvarez, M., Sanchez, S., Villalobos, J.G., Swinger, C., Schaar-Buis, D. (2019, February). *Education and training in forensic document examination: A discussion of issues and* *ideas.* Paper presented at the 2019 Annual Meeting of the American Academy of Forensic Sciences, Baltimore, MD.

Merlino, M.L., (October 2017). *Cognitive Human Factors and Sources of Bias: Sensation, Perception, and Interpretation in Decision Making Processes.* Workshop presented at the annual meeting of the European Network of Forensic Handwriting Examiners, Bled, Slovenia.

Merlino, M.L., (September 2017). *Cognitive Human Factors and Sources of Bias: Sensation, Perception, and Interpretation in Decision Making Processes.* Workshop presented at the Scottish Police Authority Crime Campus, Gartcosh, Scotland.

Merlino, M.L. (February 2017). *Cognitive Human factors and the Social Construction of Forensic Expertise, Reliability, and Validity in the Post-*Daubert *Era.* Paper presented at the symposium titled *The Evolution of* Daubert *and Its Effects on the Forensic Sciences*, 2017 Annual Scientific Meeting of the American Academy of Forensic Sciences, New Orleans, LA.

Merlino, M.L., Freeman, T.M., Springer, V., Dahir, V.B., Hammond, D., Dyer, A., & Found, B.J. (October 2016). *Cognitive Human Factors and Sources of Bias: Sensation, Perception, and Interpretation in Decision Making Processes.* Paper presented at the 2016 Annual Meeting of the Association of Forensic Document Examiners, Myrtle Beach, SC.

Merlino, M.L., Dahir, V.B., Hammond, D., Dyer, A., & Found, B.J. (April 2016). *Cognitive Human Factors and Forensic Document Examiner Methods and Procedures.* Paper presented at the Annual Meeting of the Southeastern Association of Forensic Document Examiners, Duluth, GA.