Journal of the American Society of Questioned Document Examiners, Inc.
Volume 19, Number 1, June 2016

A Blind Study on the Reliability of Hand Printing Identification by Forensic Document Examiners

Linda L. Mitchell, D-ABFDE[1] and Mara Merlino, Ph.D.[2]

[1]Forensic QDE Lab, LLC, 243 S. Escondido Blvd. #550, Escondido, CA 92025-4116
[2]Associate Professor of Psychology and Sociology; Coordinator, MA Program in Interdisciplinary Behavioral Science; Division of Behavioral and Social Sciences, Kentucky State University

Recent court cases involving that hand printing, specifically block letters, have surfaced in the news and among the forensic document examiner community. (United States of America v. Gerald Johnsted, 2013, Florida v. Jesse Lee Miller, 2014) Moreover, the current trend in elementary education is away from cursive instruction so it is likely that printing will become even more commonly used in every day writing. It is very frequently seen in forms, addresses and work-related communications as well. As a result, it has become increasingly important to verify for the purposes of a Daubert Challenge that forensic examination of hand printed documents by qualified Forensic Document Examiners (FDE) is reliable.

This study was undertaken in partial response to a court opinion, citing a lack of supporting research in the area of hand printing. It is the purpose of this research to test two hypotheses:

The qualified FDE can very reliably identify block letter hand printing (ALL CAP).

The qualified FDE can do so using the same methods and protocols as in the identification of cursive handwriting.

Each of fifty-three (53) qualified FDE participants was provided with a study packet intended to reflect evidence and procedures similar to normal bench work in a condensed form. Twenty-five (25) questioned writings were offered in individual "case packets." Each case included sample writings of three potential "suspects." Participants were asked to opine (on a nine-point scale) whether or not each suspect was the writer.

The compiled data clearly supports both of the hypotheses; (1) that qualified FDE's reliability rate with only 2.28% of calls inaccurate, when identifying (or excluding) a writer of block hand printing, and (2) that qualified FDE's do so using the same methods and protocols as in the identification of cursive handwriting.

Introduction

The reliability of opinions offered by forensic document examiners has come into question since the early days of handwriting identification and the writings of Albert S. Osborn.[1,2,3] Every challenge has compelled professionals in the field to study and report on the methodology and reproducibility of the protocols used to identify writers. These protocols are a part of the rigorous training required to become a forensic document examiner.

Until recent years, it has been broadly understood and accepted as a part of basic FDE training that analysis has been applicable to all handwritten documents.[4] With the advent of the Daubert decision,[5] judges acting as gatekeepers are charged with the responsibility to verify the methodology and supporting scientific research applied by the expert witness. The NAS Report on Forensic Science[6] has further challenged the FDE profession to test for and assure reproducible results when FDE's apply their training and skills to achieving opinions.

The authors listed above take full responsibility for the design, application, content, interpretation of data, and production of this paper. Research was conducted from the offices of the authors and was not funded by any private or public entity.

This study was conceived and developed in answer to a specific Daubert Challenge but the results will be applicable to subsequent challenges, when and if they occur. The question was… Is there research supporting the hypothesis that a properly trained forensic document examiner has the skill to reliably identify a writer by analyzing his or her block hand printing? This question was posed by Judge William M. Conley, in the 7th US Federal Court for the Western District of Wisconsin. After hearing testimony and supporting evidence, Judge Conley determined that "there had not been compelling research presented supporting the premise that hand printing (specifically) can be reliably identified using the same protocol as is used to identify cursive handwriting." (United States of America v. Gerald Johnsted, 2013)

Not long thereafter, in May of 2014, a similar Daubert Challenge was brought in the 15th District Circuit Court (Florida). At this proceeding, the forensic document examiner represented the act of handwriting as a mechanical process which is analyzed on that basis. As a result, the presiding Judge determined that "hand printing is not a bar on admissibility [of testimony.]" (Florida v. Jesse Lee Miller, 2014)

Had there been specific research supporting the question of hand printing identification, experts in both cases would have had a response to the question of skills and abilities of qualified forensic document examiners in identifying hand printing. This paper represents the results of such a research study.

Methods and Materials

Subjects

In this study the authors intended to provide normal "course of business"-type hand printing for analysis, rather than interject the variables of disguise or simulation. Consequently, the subjects who provided hand printing for the test consisted of writers who commonly print in block letters (all capital letters, upper and/or lower) or who use block letters in their work environment. This population would be expected to have a more natural speed and line quality, thus minimizing the potential for self-conscious drawing or slowness of execution. Writers in this group included military personnel, engineers, and law enforcement among others.

Each writer provided a short demographic profile consisting of their age, sex, handedness, occupation and normal writing or printing habits. They were asked to copy the London Letter onto one lined page and on a second lined page write extemporaneously about themselves, their aspirations, their family or anything else that came to mind. Each writer's product was assigned an identifying number; names of the writers were not requested. For the purposes of this study, The London Letter would be used as the "Requested Writing" exemplar and a short snippet of the extemporaneous writing would be supplied to the participant as the "Writing in Question."

Participants

The 53 participants in this study are practicing FDE's, trained according to the guidelines set forth by the Scientific Working Group for Forensic Document Examiners (SWGDOC.org), and have been vetted by one or more of several FDE professional organizations.[7] There was no remuneration offered. Each FDE participant was given an identifying number between 001 and 150. All data received and analyzed was anonymous from that point forward. Only the authors had access to the roster of participants.

Study Materials

Once all of the collected sample writing was juxtaposed, it was determined that certain of the hand printing clearly shared similar pictorial qualities. These samples were gathered in groups and ultimately used as common components of case packet questions. The hand printing samples that were not included as part of this study were either not written in all block letters or were so individualized in their execution they would be considered fundamentally different than the other collected samples. Selecting the samples that were included was a subjective process on the part of the author; it was intended to provide a scenario more closely resembling bench work wherein a forensic document examiner (FDE) would be required to provide an expert opinion.

Twenty-five (25) case packets were offered for analysis in this study. For each packet, a two-line segment was extracted from the spontaneous hand printed sample and designated as the writing in question. Three London Letter samples were selected from the same pictorial group and designated as requested writing, ostensibly from

three possible suspects. In several instances the true writer's London Letter was not included among the choices. FDE participants were asked to opine as to authorship of the questioned writing for each of the three potential writers using the currently accepted 9-point scale.[8]

Study materials were provided on a DVD containing all of the packet materials. For some participants, the study materials were provided as internet transmissions. Answers were collected online using the Adobe Form Central service[9]. For each packet group, participants were to select one of nine (9) radio buttons representing their opinion and level of confidence for each Known writer as the being the author of the questioned sample in that group. FDE's were allowed to leave and return to their online answer sheet as necessary over a six-week period (or less depending on when the study was received).

Results

Of the 1,325 total possible calls (25 questioned samples x 3 calls per sample—75 calls per examiner x 53 examiners—1,325 total calls), 1,322 were made by the examiners. Of the 1,322 calls made, 1,216 calls (91.98%) were made correctly (e.g., the writer of the questioned sample was either correctly identified from among the three known samples or writers of the three known samples were correctly eliminated as the writer of the questioned sample). Thirty-one (2.28%) of the 1,322 calls were incorrect. Seventy-five (5.6%) of the calls were inconclusive. Three calls were non-responses. The characteristics of the questioned printing samples are presented in Table 1.

Table 1. Questioned Hand Printing Sample Information Content, Number of Correct Calls*, and Mean Call Examiner Call Confidence.

*Sample***	*Words*	*Characters*	*Allographs*	*Word %*	*Character %*	*Allograph %*	*# Calls Correct*	*Mean Call Confidence*
Q1(G)	17	90	74	20.0%	19.8%	20.0%	53	1.87
Q2(G)	13	72	60	15.3%	15.9%	16.2%	49	2.18
Q3(G)	18	95	72	21.2%	20.9%	19.5%	52	2.56
Q4(G)	20	107	88	23.5%	23.6%	23.8%	52	2.09
Q5(G)	18	86	69	21.2%	18.9%	18.6%	48	2.65
Q6(G)	18	96	79	21.2%	21.1%	21.4%	51	1.79
Q7(G)	17	96	80	20.0%	21.1%	21.6%	49	2.16
Q8(N)	21	112	92	24.7%	24.7%	24.9%	46	2.51
Q9(G)	11	49	39	12.9%	10.8%	10.5%	52	1.82
Q10(N)	20	91	72	23.5%	20.0%	19.5%	51	2.51
Q11(G)	16	79	64	18.8%	17.4%	17.3%	52	1.75
Q12(G)	4	21	18	4.7%	4.6%	4.9%	34	2.89
Q13(G)	23	102	80	27.1%	22.5%	21.6%	51	2.16
Q14(N)	21	111	91	24.7%	24.4%	24.6%	46	2.68
Q15(G)	18	85	68	21.2%	18.7%	18.4%	50	2.26
Q16(G)	20	85	66	23.5%	18.7%	17.8%	50	2.05
Q17(G)	22	123	102	25.9%	27.1%	27.6%	50	2.18
Q18(N)	16	82	67	18.8%	18.1%	18.1%	52	2.64
Q19(N)	18	117	100	21.2%	25.8%	27.0%	51	2.67
Q20(G)	13	77	65	15.3%	17.0%	17.6%	38	2.47
Q21(G)	16	82	67	18.8%	18.1%	18.1%	53	1.86
Q22(G)	13	69	57	15.3%	15.2%	15.4%	53	2.34
Q23(N)	14	98	85	16.5%	21.6%	23.0%	51	2.56
Q24(G)	23	105	83	27.1%	23.1%	22.4%	52	2.02
Q25(G)	15	79	65	17.6%	17.4%	17.6%	47	2.04

*Correct calls are defined as a correct determination of whether or not the questioned signature is written by one of the writers of the three known specimens.
**G = genuine questioned specimen; N = questioned specimen not represented among known samples.

"Words" are defined as the total number of words in the questioned sample. "Characters" are defined as the total number of letter forms and spaces in the questioned sample. "Allographs" are defined as the total number of letter forms (excluding spaces) in the questioned sample. These informational characteristics were obtained using the word count function of MS Word. Word, character, and allograph percentages were determined by dividing the number of words, characters, and allographs in the sample by the corresponding characteristics of the London Letter (85 words; 454 characters; 370 allographs).

Call certainty was measured using the nine-level bipolar "authorship" scale ranging from 1 (elimination) to 9 (identification), with the point "5" representing the "inconclusive" position. The scale values were reverse-coded to create a unipolar measure with five levels: 0—inconclusive; 1—indications did/did not write; 2—probably did/did not write; 3—strong probability did/did not write; and 4—elimination/identification. This allowed us to analyze and compare the certainty for eliminations to that of identifications.

A multiple regression was conducted to determine whether there was a relationship between the information content (percentage of London Letter characters, percentage of London Letter letters, and percentage of London Letter allographs) contained in the questioned samples and the mean examiner call certainty for each set of three knowns. The model was not statistically significant ($p = .238$), indicating that there was no relationship between the amount of information available to the examiner in the questioned printing sample and the examiner's call confidence.

Multivariate analyses of variance (MANOVAs) were performed to investigate whether there were any differences in examiners' reported certainty according to the length of time they had been performing examinations. Significant differences in authorship confidence opinions were found in 11 of the 25 hand printing comparison task samples. Table 2 presents the *post hoc* analyses for the statistically significant outcomes.

Table 2. Mean Examiner Call Confidence by Years of Examiner Experience and Hand Printing Specimen.

		Hand Printing Sample					
		K1A (Correct)		K1B		K1C	
Years of Experience	N	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
< 10	16	2.94	0.23	2.63	0.25	2.75	0.193
10 to 20	14	3	0.23	3.36	0.24	3.3	0.19
> 20	23	2.75	0.21	3.29	0.22	3.43* ($p = .040$)	0.171
		K4A		K4B (Correct)		K4C	
	N	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
< 10	16	2.19	0.286	2.31	0.262	2.5	0.21
10 to 20	14	2.93	0.305	3.14	0.28	3.43* ($p = .012$)	0.225
> 20	22	3.14* ($p = .044$)	0.244	3.14	0.223	3.18	0.179
		K6A (Correct)		K6B		K6C	
	N	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
< 10	16	2.75	0.263	2.63	0.221	2.63	0.23
10 to 20	14	3.36	0.281	3.5* ($p = .028$)	0.236	3.36	0.246
> 20	23	3.35	0.22	3.43* ($p = .021$)	0.184	3.39* ($p = .041$)	0.192
		K7A		K7B (Correct)		K7C	
	N	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
< 10	16	2.31	0.229	2.25	0.237	2.38	0.206
10 to 20	14	3	0.245	3	0.253	3.21* ($p = .023$)	0.221
> 20	22	3.09* ($p = .039$)	0.196	2.95	0.202	3.18* ($p = .014$)	0.176
		K9A		K9B		K9C (Correct)	
	N	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
< 10	16	2.56	0.215	2.69	0.223	2.88	0.207

10 to 20	14	3.5* (p = .013)	0.23	3.43	0.238	3.5	0.221
> 20	23	3.39* (p = .014)	0.179	3.3	0.186	3.3	0.172
		K12A (Correct)		**K12B**		**K12C**	
	N	***M***	***SEM***	***M***	***SEM***	***M***	***SEM***
< 10	16	1	0.278	1.25	0.299	1.13	0.3
10 to 20	14	1.21	0.298	1.5	0.319	1.5	0.321
> 20	23	1.57	0.232	2.3* (p = .028)	0.249	2.3* (p = .012)	0.25
		K14A* (Correct)		**K14B**		**K14C**	
	N	***M***	***SEM***	***M***	***SEM***	***M***	***SEM***
< 10	16	1.44	0.267	1.81	0.259	1.63	0.262
10 to 20	14	2.36	0.285	2.5	0.277	2.43	0.28
> 20	23	2.52* (p = .009)	0.222	2.48	0.216	2.48* (p = .047)	0.219
		K16A		**K16B* (Correct)**		**K16C**	
	N	***M***	***SEM***	***M***	***SEM***	***M***	***SEM***
< 10	16	2.5	0.255	2.63	0.271	2.44	0.238
10 to 20	14	3	0.272	2.64	0.29	3.07	0.254
× 20	23	3.22	0.212	3	0.226	3.26* (p = .032)	0.198
		K17A		**K17B**		**K17C* (Correct)**	
	N	***M***	***SEM***	***M***	***SEM***	***M***	***SEM***
< 10	16	2.19	0.265	2.19	0.271	2	0.297
10 to 20	14	2.86	0.283	2.86	0.289	2.57	0.318
> 20	23	3.22* (p = .013)	0.221	3.26* (p = .011)	0.226	2.65	0.248
		K24A		**K24B**		**K24C* (Correct)**	
	N	***M***	***SEM***	***M***	***SEM***	***M***	***SEM***
< 10	16	2.44	0.234	2.56	0.258	2.5	0.261
10 to 20	14	3.43* (p = .017)	0.251	3.29	0.276	3.14	0.279
> 20	23	3.13	0.196	3.26	0.215	2.96	0.218
		K25A		**K25B**		**K25C* (Correct)**	
	N	***M***	***SEM***	***M***	***SEM***	***M***	***SEM***
< 10	16	2.38	0.244	2.19	0.275	2.38	0.318
10 to 20	14	3.36* (p = .025)	0.261	3.07	0.294	3.14	0.34
> 20	23	3.13	0.203	3.04	0.229	2.57	0.266

*Significantly different from examiners with fewer than 10 years of experience.

Although overall call accuracy was very high (91.98% correct), results of these analyses indicated that in the 11 instances where significant differences in confidence were identified, on average those examiners with greater than ten years of experience were significantly more confident in their decisions than were examiners with fewer than 10 years of experience.

Geographical Differences

No significant differences in call accuracy or confidence were identified between examiners from within the United States and examiners from other countries.

Discussion

The results of this study illustrate the ability of forensic document examiners to identify writers of hand printing using the same techniques and protocols as are used to identify cursive writing[10]. In the early days of Forensic Document Examination hand printing, referred to as Pen-Printing by Albert S Osborn[11], discussed class characteristics and peculiarities in his *Questioned Documents*

publication and dedicates an entire chapter to the topic in the second edition, incorporating it into the examination process. In the following century authorities such as Wilson Harrison, Ordway Hilton, Roy Huber, AM Headrick, Jan Kelly and Brian Lindblom have elaborated, clarified and redefined the analysis of hand printing,[12,13,14,15] including it as part of the analysis of handwriting in general. Modern characteristics of hand printing were further defined and illustrated in the 1999 paper of the same name by Ellen Schuetzner.[16] And finally, Ron Morris defined "Five Principals of Handwriting and Hand Printing Identification" in his 2005 paper, as further confirmation that the principles of analysis have always and still apply to both hand printing and cursive handwriting.

If hand printing is considered part of the FDE's analytical expertise, then is there any difference in the identification skills between printing and cursive? At the 2003 ASQDE Annual Conference of ASQDE, Moshe Kam, et al. examined this topic as part of a compendium of their research of FDE skills and reliability. His data reflected an accuracy rate for FDE in identifying hand printing and cursive on par with a layperson, but incorrect associations were made four times more often by laypersons than by FDE's, supporting the proposition that the professional opinion is more reliable. It is significant to note that in this study, the error rate of the FDE participants is 2.28%.

Study Limitations

As with any study such as this, questions arise as to the validity of the content, structure of the materials, and the execution of data collection. These limitations may include the number of FDE's who participated, the difference in performance between those in real life versus those in a structured environment such as this, and the limited amount of hand printing samples for examination. It can also be conjectured that FDE's are less cautious in a professional study scenario than in actual case work when determining the level of confidence for their opinions.

Given these potential targets of discussion, we can state that the opinions on either side of neutral—to eliminate or identify—represent correct or incorrect answers despite the level of certainty. These answers accurately reflect the ability of the FDE's under the given conditions and can be measured as to ground truth. The number of participants represents a statistically significant percentage (approximately 15-20%) of the potential FDE pool worldwide.[17]

In the interest of full disclosure—The study materials were originally designed to include only a few words as the questioned hand printing in each packet. This design was later changed to include more questioned writing. The logic there was to avoid inconclusive answers due to the lack of adequate writing samples. In the process of transmitting study materials via email, some of the participants received the case packets in their original form. When the mistake was recognized, each of those participants was offered the opportunity to receive the updated packet and resubmit their answers. Of those FDE's who elected to receive the updated packets, none chose to change their answers.

Future Directions

In keeping with such studies as those mentioned above, the authors intend to expand this study to include participation by untrained adults. The same testing format will be applied; however, it may become necessary to offer lay participants an incentive to provide thoughtful answers and complete all 25 questions. Additionally, it will be necessary to decide the extent of explanation that will be necessary to define and differentiate the nine levels of confidence. Other potential studies could include upper-and lowercase hand printing and require that a new set of samples be collected and formatted similar to this study.

These studies may also become a springboard for eye-tracking evaluations which would be designed to evaluate the human factors involved in the decision-making process.[18]

Conclusion

In this study of forensic document examiners applying the training and laboratory procedures that conform to SWGDOG.org standards, we have assessed an error rate of 2.28% when identifying a writer of block hand printing. Further, we would expect that the real-life results of such examinations would reflect a better performance by the handwriting expert in the identification process. We also recognize that no single or series of research tests can predict the accuracy rate in any given situation. Still in all, the data from this test

provides a basis for judges and others to evaluate the general skill set of forensic document examiners (FDE) to identify block hand printing.

Acknowledgments:

Gina Hunter, FDE trainee, San Diego Sheriff Crime Lab; Marie Durina, D-ABFDE, former senior FDE, San Diego Sheriff Crime Lab

References

1. Osborn, Albert S. Questioned Documents, Second Edition, (p374), Nelson-Hall Co., 1929

2. Notes of Advisory Committee on Rules (1972) http://federalevidence.com/advisory-committee-notes/FRE.702

3. United States v. Prime, 431 F.3d 1147 (9th Cir. 2005)

4. SWGDOC Standard for Examination of Handwritten Items. http://swgdoc.org/index.php/standards/published-standards

5. http://www.justice.gov/ust/kuhmo-tire-and-admissibility-expert-testimony

6. National Research Council. Strengthening Forensic Science in the United States: A Path Forward. Washington, DC: The National Academies Press, 2009.

7. AAFS—American Academy of Forensic Science (QD Section), ABFDE – American Board of Forensic Document Examiners, ASQDE – American Society of Questioned Document Examiners, MAAFS—Mid-Atlantic Association of Forensic Scientists, SWAFDE—Southwestern Association of Forensic Document Examiners, and ASFDE – Australasian Society of Forensic Document Examiners

8. While not all FDE's use the nine-point scale and opt for a more abbreviated version, this scale provides the broadest range of levels of confidence choices for those who are accustomed to using it.

9. This service has been discontinued.

10. Of the 53 participants, 52 indicated that they use the same protocol to identify both hand printing and cursive.

11. Osborn, Albert S. *Questioned Documents, A Study of Questioned Documents with an Outline of Methods by Which the Fact May Be Discovered and Shown.* The Lawyers' Co-operative Publishing Co., 1910.

12. Harrison, Wilson R. *Suspect Documents, Their Scientific Examination*. Frederick A. Praeger, Publishers, 1958.

13. Hilton, Ordway *Scientific Examination of Questioned Documents, Revised Edition.* CRC Press, LLC, 1993.

14. Huber, Roy & Headrick, A.M. *Handwriting Identification: Facts and Fundamentals*. CRC Press, 1999.

15. Kelly, Jan Seaman & Lindblom, Brian S. *Scientific Examination of Questioned Documents, Second Edition.* Taylor & Francis Group, 2006

16. Schuetzner, Ellen Mulcrone. *Class Characteristics of Hand Printing,* The American Society of Questioned Document Examiners, Inc. 1999.

17. As specified in our qualification for study participants.

18. Merlino, M.L., Freeman, T.M., Dahir, B., Springer, V., Hammond, D., Dyer, A.D., & Found, B.J. (2012). *Reliability and validity in forensic document examination: A demonstration and discussion of eye-tracking output*. Paper presented at the Annual General Meetig of the American Society of Questioned Document Examiners, Charleston, SC. Abstract retrieved from http://www.asqde.org/AGM/previous_meetings/programs_and_abstracts/2012Abstracts.pdf.