J Forensic Sci, November 2003, Vol. 48, No. 6
Paper ID JFS2002321_486
Available online at: www.astm.org

Moshe Kam,[1] *Ph.D. and Erwei Lin,*[1] *M.S.*

Writer Identification Using Hand-Printed and Non-Hand-Printed Questioned Documents

ABSTRACT: Several federal district court judges have recently referred to the purported lack of information on the proficiency of forensic document examiners (FDEs) in identifying writers of hand-printed documents. In order to provide the necessary information, we have re-analyzed data on writer identification that were collected in 1996 from 90 forensic document examiners and 34 laypersons. These data were used previously to assess the proficiency of FDEs using handwritten documents in several different types of writing. In the new analysis we separated data on hand-printed (HP) documents from data on non-hand-printed (NHP) documents and compiled error rates and statistics in each category. The main findings are: (1) whether or not the documents were hand-printed, the performance of FDEs was much better than that of laypersons; (2) statistical tests found no difference between the data provided by the FDEs in the HP and NHP categories; (3) statistical tests found no difference between the data provided by laypersons in the HP and NHP categories; and (4) statistical tests found differences between the data provided by the FDEs and the laypersons in both the HP and NHP categories. Similar results were obtained when hand-printed documents were compared to cursive documents and when cursive documents were compared to non-cursive documents. All the evidence indicates that in our proficiency test the performance of FDEs in writer identification was much better than the performance of laypersons in each one of the following document categories: (1) hand-printed; (2) non-hand-printed; (3) cursive; and (4) non-cursive.

KEYWORDS: forensic science, forensic document examination, questioned documents, proficiency testing, writer identification, hand-printed documents, script handwriting, cursive handwriting

The proficiency of forensic document examiners (FDEs) was the subject of several controlled studies conducted in the last decade (1–6). These tests examined the performance of FDEs and laypersons in handwriting identification (1–3) and in authentication of signatures (4–6). Recently, however, several federal district court judges have required information on the abilities of FDEs in the specific area of hand-printed documents. In three cases (7) the purported lack of such information was a factor in the court's ruling. To address this need, we provide here a re-analysis of the data from the 1996 controlled proficiency test (reported originally in Ref 2). This re-analysis treats hand-printed (HP) and non-hand-printed (NHP) documents separately. In addition, we provide information on FDE and layperson writer identification performance using cursive (C) and non-cursive (NC) documents. While the original motivation for this study was the HP vs. NHP comparison, we provide also HP vs. C, and C vs. NC comparisons.

General Description of the 1996 Proficiency Test and Its Subjects

The professionals who took the 1996 tests were either employed or recently retired FDEs, employed by law enforcement agencies or in for-profit private practice. Almost all of these FDEs were certified by, or were members of, one or more of the following organizations: American Academy of Forensic Sciences—Questioned Documents Section; American Board of Forensic Document Examiners; Southeastern Association of Forensic Document Examiners; Southwestern Association of Forensic Document Examiners; and the American Society of Questioned Document Examiners. Members of the Northeastern group were individually invited to the exam using lists of examiners residing in the vicinity of New York and Washington, D.C. who are members of these professional organizations. Members of the Southwestern and Southeastern groups were attendees of the May 1996 professional meetings of FDE regional associations.

The layperson test-takers were students and educators from the Greater Philadelphia area, about four-fifths of them holders of college degrees (B.A., B.Sc., M.A., M.Sc., M.Ed., M.B.A., Ph.D.) in Engineering, Education, or Management. The rest were senior-class undergraduate students in Engineering. The laypersons were screened for education level in order to match the educational profile of the professional groups. For more information about the test-takers and about monetary incentives to laypersons, see Refs 2 and 3.

Each test-taker (FDE or layperson) was given an "unknown" package (six original handwritten documents, not necessarily by the same writer), and a "database" package (24 original handwritten documents, not necessarily by the same writer). The test-taker was asked to match each one of the documents in the unknown package to all the documents in the database package that were written by the same person. A "match" was declared if the test-taker was able to find an association that conforms to the definition for Identification or for Strong Probability in Standard E 1658 of the American Society for Testing and Materials (ASTM) (8). In Ref 2 we calculated error rates for both FDE and layperson test-takers and performed statistical tests on these data.

Methods

For additional information on the subjects, materials, and procedures used in the 1996 test, please see Ref 2. In preparation for the

[1] Data Fusion Laboratory, Electrical and Computer Engineering Department, Drexel University, 3141 Chestnut Street, Philadelphia, PA.

Received 11 Sept. 2002; and in revised form 13 Jan. 2003 and 12 May 2003; accepted 18 May 2003; published 27 Aug. 2003.

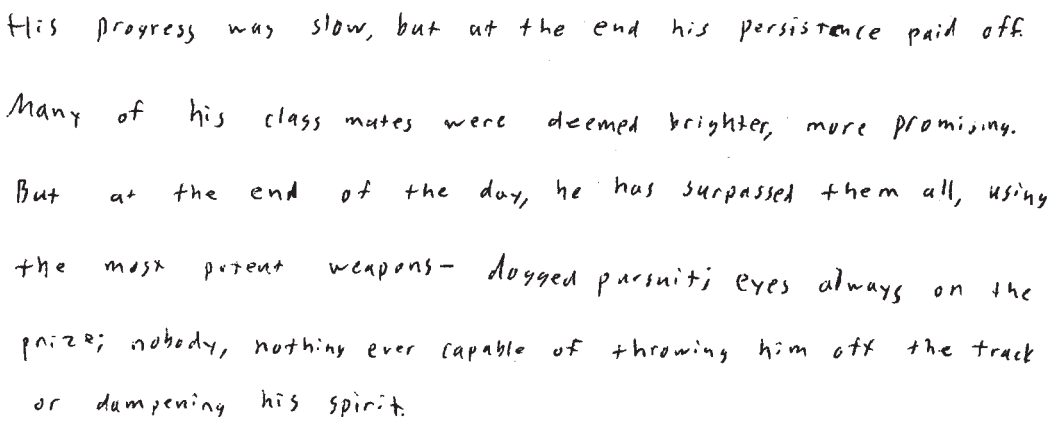

FIG. 1—*A hand-printed writing sample from the 1996 test.*

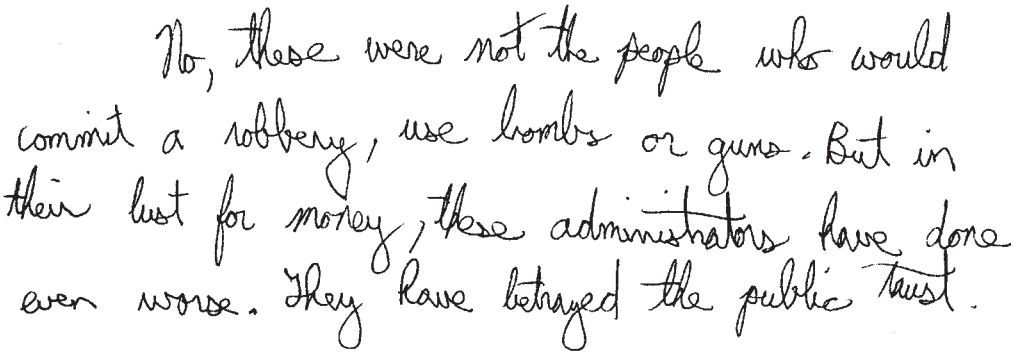

FIG. 2—*A cursive writing sample from the 1996 test.*

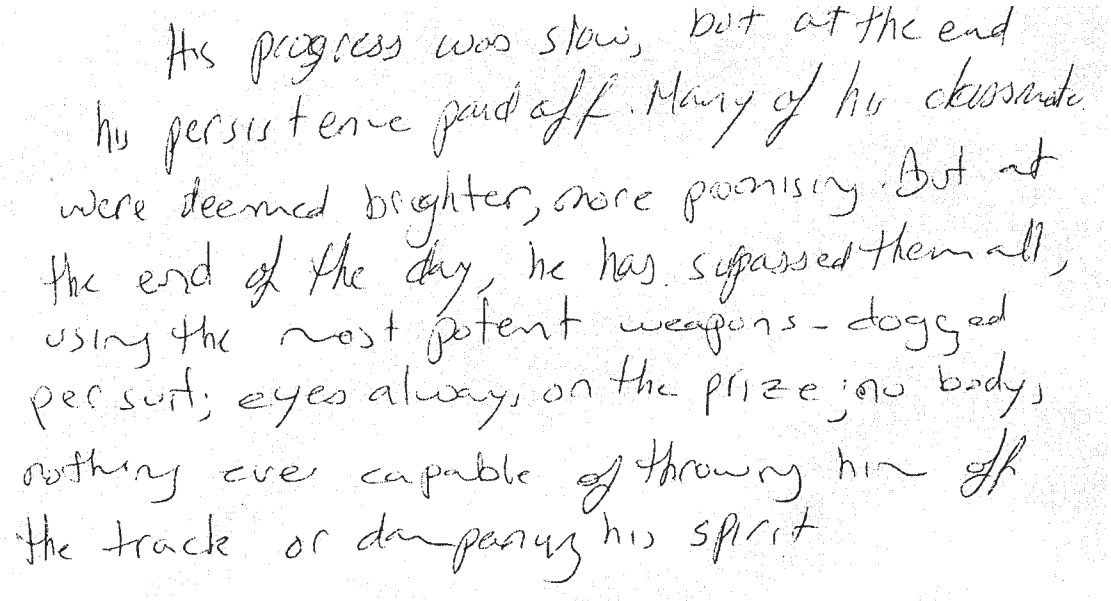

FIG. 3—*Mixed writing sample from the 1996 test.*

present task, 55 of the 60 unknown document packets used in the 1996 test were classified according to the type of writing (for a total of 330 documents out of the 360 used in the 1996 test). Each document was classified as hand-printed, cursive, or mixed. Hand-printed and cursive samples from the 1996 test were identified based upon their respective definitions in ASTM standard E 2195 (11). A sample was classified as mixed if it contained significant subsets of handwriting of both hand-printing and cursive types.

Examples of cursive and hand-printing writing copybooks are available in many textbooks, with the Palmer and Zaner-Bloser systems cited as the most popular in the United States (e.g., Refs 9 and 10). Figures 1–3 show samples that were used in the actual

tests: the sample in Fig. 1 is hand-printed, the sample in Fig. 2 is cursive, and the sample in Fig. 3 is mixed. We refer to the 330 documents used in the re-analysis as the "test sample." We refer to cursive and mixed documents collectively as non-hand-printed. We refer to the hand-printed and mixed documents collectively as non-cursive.

The 1996 answer sheets were tabulated *de novo*, independently of any previous studies and statistics. The analysis followed the criteria set forth in Ref 2. As in the previous analysis, each unknown (or "questioned") document was considered to have been matched to a database document if the test-taker declared that the two documents were associated to the degree of Identification or Strong Probability as defined in the E 1658 standard.

Two types of error could have occurred:

1. A questioned document that had a matching document in the database was not associated with that matching document by the test-taker.
2. A questioned document was erroneously associated by the test-taker with a database document that did not match it.

The first kind of error (1) would lower the group hit rate (GHR). The second kind of error (2) would increase the wrong association rate (WAR).

Definition 1—the group hit rate (GHR) is the number of matches declared correctly by members of the tested group divided by the number of correct matches actually present in that group's tests.

Ideally the GHR is one (100%). In the 1996 tests, the FDEs had a GHR of 87.9% and the laypersons had a GHR of 87.7%.

TABLE 1—*Data on the documents in the test sample.*

Group	Document Type	Number of Questioned Documents Examined by Group	Number of True Matches in Group's Tests
FDEs (90)	All	536	397
	Hand-printed	247	182
	Non-hand-printed	289	215
	Cursive	235	176
	Non-cursive	301	221
Laypersons (34)	All	204	138
	Hand-printed	89	65
	Non-hand-printed	115	73
	Cursive	91	59
	Non-cursive	113	79

Definition 2—the wrong association rate (WAR) is the number of questioned documents that were matched erroneously by the members of the tested group divided by the total number of questioned documents that existed in that group's tests.

Ideally the WAR is zero (0%). In the 1996 tests, the FDEs had a WAR of 6.5%, while the laypersons had a WAR of 38.3%.

Table 1 provides information about the test sample. It shows (1) the number of documents of each type analyzed by each group of test-takers, and (2) the number of correct associations actually present in the tests taken by each group for each type.

Performance Results

Table 2 provides the GHR and WAR for the FDEs and the laypersons. These are given for the ideal case, the 1996 test documents, and the following subsets of the test sample: (1) all documents; (2) hand-printed documents; (3) non-hand-printed documents; (4) cursive documents; and (5) non-cursive documents.

It is apparent from Table 2 that the FDEs performed much better than the laypersons on all document types. This is evident when one compares the WAR values of the test-takers in each document category (the GHRs are comparable).

As we explain below, the category percentage fluctuations in Table 2 within each column (Rows 5–8) are *not* statistically significant. On the other hand, the differences between the WAR values of the FDEs (Column 3, Rows 5–8) and the laypersons (Column 5, Rows 5–8) *are* significant.

Statistical Significance

We compared the distributions of the GHR and WAR in the HP and NHP categories within each of the test-taker groups and between the test-taker groups. To this end we used the Kolmogorov-Smirnov (KS) two-sample test (12,13). We also provide results from the rank test of Mann and Whitney (MW) (13). Our threshold for the rejection of an hypothesis was $p = 0.05$ with the KS test[2].

Table 3 shows the results of hypothesis testing for hand-printed versus non-hand-printed documents. The first line in this table reads: "should we reject the hypothesis that the *group hit rate* samples collected from the *FDEs* for *hand-printed documents* and the (group hit rate) samples collected from *FDEs* for *non-hand-printed documents* came from the same population?" The answer is "do not reject the hypothesis." The table also provides the values of the

[2] We used the functions Kolmogorov Smirnov Z and Mann Whitney U (non-parametric tests—two independent samples) in the popular software package SPSS for Windows, release 11.0.1.

TABLE 2—*Group hit rates and wrong association rates.*

Condition	Forensic Document Examiners		Laypersons	
	Group Hit Rate (GHR)	Wrong Association Rate (WAR)	Group Hit Rate (GHR)	Wrong Association Rate (WAR)
Ideal	100%	0%	100%	0%
1996 test	87.9%	6.5%	87.7%	38.3%
Test sample, all documents	87.15%	7.3%	89.8%	39.2%
Test sample, hand-printed documents only	88.5%	9.3%	93.85%	40.45%
Test sample, non-hand-printed documents only	86.0%	5.5%	86.3%	38.26%
Test sample, cursive documents only	86.93%	4.7%	89.83%	42.86%
Test sample, non-cursive documents only	86.43%	8.97%	89.87%	36.28%

TABLE 3—*Hypothesis testing for hand-printed and non-hand-printed documents.*

Should we reject the hypothesis that . . .						Kolmogorov-Smirnov Z		Mann-Whitney U/ Wilcoxon W	
the	samples collected from the	for	and the samples collected from	for	came from the same population?	*Z* Value	*p* Value	*U*/*W* Value	*p* Value
GHR	FDEs	HP documents	FDEs	NHP documents	*Do not reject*	0.887	*0.411*	2659.5/5585.5	0.085
WAR	FDEs	HP documents	FDEs	NHP documents	*Do not reject*	0.671	*0.759*	3755.5/7850.5	0.092
GHR	Laypersons	HP documents	Laypersons	NHP documents	*Do not reject*	0.645	*0.799*	561.0/1156.0	0.085
WAR	Laypersons	HP documents	Laypersons	NHP documents	*Do not reject*	0.423	*0.994*	515.5/1110.5	0.599
GHR	Laypersons	HP documents	FDEs	HP documents	*Do not reject*	0.506	*0.960*	1365.0/5460	0.187
WAR	Laypersons	HP documents	FDEs	HP documents	*Reject*	1.593	*0.012*	1000.0/5095.0	0.000
GHR	Laypersons	NHP documents	FDEs	NHP documents	*Do not reject*	0.350	*1.000*	1468.5/5563.5	0.579
WAR	Laypersons	NHP documents	FDEs	NHP documents	*Reject*	2.516	*0.000*	706/4801.5	0.000

TABLE 4—*Hypothesis testing for hand-printed and cursive documents.*

Should we reject the hypothesis that . . .						Kolmogorov-Smirnov Z		Mann-Whitney U/ Wilcoxon W	
the	samples collected from the	for	and the samples collected from	for	came from the same population?	*Z* Value	*p* Value	*U*/*W* Value	*p* Value
GHR	FDEs	HP documents	FDEs	Cursive documents	*Do not reject*	0.308	*1.000*	2764.5/5392.5	0.557
WAR	FDEs	HP documents	FDEs	Cursive documents	*Do not reject*	0.731	*0.659*	3015.5/7530.5	0.039
GHR	Laypersons	HP documents	Laypersons	Cursive documents	*Do not reject*	0.321	*1.000*	368.5/746	0.334
WAR	Laypersons	HP documents	Laypersons	Cursive documents	*Do not reject*	0.492	*0.969*	481/1042.0	0.394
GHR	Laypersons	HP documents	FDEs	HP documents	*Do not reject*	0.506	*0.960*	1365.0/5460	0.187
WAR	Laypersons	HP documents	FDEs	HP documents	*Reject*	1.593	*0.012*	1000.00/5095.00	0.000
GHR	Laypersons	Cursive documents	FDEs	Cursive documents	*Do not reject*	0.328	*1.000*	913.5/3541.5	0.531
WAR	Laypersons	Cursive documents	FDEs	Cursive documents	*Reject*	2.672	*0.000*	623.5/4539.5	0.000

TABLE 5—*Hypothesis testing for cursive and non-cursive documents.*

Should we reject the hypothesis that . . .						Kolmogorov-Smirnov Z		Mann-Whitney U/ Wilcoxon W	
the	samples collected from the	for	and the samples collected from	for	came from the same population?	*Z* Value	*p* Value	*U*/*W* Value	*p* Value
GHR	FDEs	Cursive documents	FDEs	Non-cursive documents	Do not reject	0.408	0.996	8084.5/6825.5	0.956
WAR	FDEs	Cursive documents	FDEs	Non-cursive documents	Do not reject	0.728	0.665	3536/7452	0.045
GHR	Laypersons	Cursive documents	Laypersons	Non-cursive documents	Do not reject	0.159	1.000	418.0/1914	0.991
WAR	Laypersons	Cursive documents	Laypersons	Non-cursive documents	Do not reject	0.558	0.915	419/1086	0.360
GHR	Laypersons	Cursive documents	FDEs	Cursive documents	Do not reject	0.328	1.000	913.5/3541.5	0.531
WAR	Laypersons	Cursive documents	FDEs	Cursive documents	Reject	2.672	0.000	623.5/4539.5	0.000
GHR	Laypersons	Non-cursive documents	FDEs	Non-cursive documents	Do not reject	0.444	0.989	0.250	0.485
WAR	Laypersons	Non-cursive documents	FDEs	Non-cursive documents	Reject	1.750	0.004	0.188	0.000

statistics and *p*-values from the Kolmogorov-Smirnov Z and Mann-Whitney U tests.

While our primary interest was the HP vs. NHP comparisons, we also performed HP vs. C and C vs. NC comparisons. Tables 4 and 5 show the results of these additional comparisons. The results in these tables have the same pattern as Table 3. Again we conclude that the GHR and WAR fluctuation within the same test-taker group as a function of the type (Table 2, Rows 5–8) are *not* statistically significant; the WAR differences between the FDEs and the laypersons (Table 2, Columns 3 and 5) are statistically significant.

Discussion

Performance on Hand-Printed Versus Non-Hand-Printed Documents

The analysis of the test sample found that for both types of handwriting (HP and NHP) the data provided by FDEs and the data provided by laypersons were significantly different. The performance of the FDEs was much superior to that of laypersons in both categories. These two observations taken together support the claim for existence of expertise in handwriting identification by FDEs for both HP and NHP documents.

The WAR of FDEs with HP documents was 3.8% worse than their WAR with NHP documents, while their GHR was 2.5% better. However, these differences were found by the KS test to be *not* statistically significant.

Results for All Document Types

Similar conclusions to the ones obtained in the HP vs. NHP comparisons were obtained when we compared HP vs. C, and C vs. NC. Our data thus support the claim for existence of expertise in handwriting identification by FDEs for each one of the handwriting types separately (hand-printed, non-hand-printed, cursive, and non-cursive), as well as in the joint category of "all documents." For each one of the document types, the performance of the FDEs was far superior to that of laypersons, and the data supplied by FDEs were significantly different than those of the laypersons.

Conclusions

We have re-analyzed the data collected for the 1996 proficiency test (2) in order to study separately the performance of test-takers on hand-printed (HP) and non-hand-printed (NHP) documents. For each group of test-takers separately, the Kolmogorov-Smirnov test did not detect statistically significant differences between data on HP documents and data on NHP documents. However, within each document type the data provided by FDEs were found to be significantly different from the data provided by laypersons. Moreover, the performance of FDEs was found to be much better than the performance of laypersons for both hand-printed and non-hand-printed documents.

Similar results were obtained when cursive vs. non-cursive and hand-printed vs. cursive comparisons were made. Data supplied by each test-taker group for any pair of document types (e.g., data supplied by FDEs on cursive and non-cursive documents) were *not* found by the KS test to be statistically different. However, when data from different test-taker groups were compared within the same document type, the data provided by FDEs and the data provided by laypersons were found to be significantly statistically different (specifically, statistical differences were found in WAR data). Moreover, FDEs had much better performance than the laypersons in each one of the handwriting types.

In all document types, laypersons were distinguished from FDEs by their tendency to over-associate documents to each other, resulting in the attribution of many questioned documents to individuals who did not write them (this resulted in similar GHR values for the two groups, but much higher WAR for laypersons).

Acknowledgment

This study was funded by Drexel University's Distinguished Professor Award and by the National Institute of Justice Award Number 2002-LT-BX-K008: Proficiency and Procedure Validation for Forensic Document Examiners.

References

1. Kam M, Wetstein J, Conn R. Proficiency of professional document examiners in writer identification. J Forensic Sci 1994;39(1):5–14.
2. Kam M, Fielding G, Conn R. Writer identification by professional document examiners. J Forensic Sci 1997;42(5):778–86.
3. Kam M, Fielding G, Conn R. The effect of monetary incentives on document examination by nonprofessionals. J Forensic Sci 1998;43(5): 1000–4.
4. Found B, Sida J, Rogers D. The development of a program for characterising forensic handwriting examiners' expertise: signature examination pilot study. Journal of Forensic Document Examination 1999; 12:69–80.
5. Kam M, Gummadidala K, Fielding G, Conn R. Signature authentication by forensic document examiners. J Forensic Sci 2001;46(4):884–8.
6. Sida J, Found B, Rogers DK. Forensic handwriting examiners' expertise for signature comparison. J Forensic Sci 2002;47(5):1–8.
7. U.S. vs. Chan Ian Salee, United States District Court for the District of Alaska, No. A0 1-0084 CR (HRH); U.S. vs. Masao Fujii, United States District Court N.D. Illinois, Eastern Division No. 00 CR 17; US vs. Edward Lee Lewis, United States District Court for the Southern District of West Virginia, No. 2:02-00042 CR.
8. ASTM International. ASTM standard E 1658: Standard Terminology for Expressing Conclusions of Forensic Document Examiners.
9. Freeman FN, the Zaner-Bloser Co. Handwriting aid for primary teachers: general and specific guide for grades 1–3. Columbus, OH: The Zaner-Bloser Co., 1948.
10. Kittle R. Handwriting for students of English as a second language. New York: American Book Company, 1966.
11. ASTM International. ASTM standard E2195: Terminology Relating to the Examination of Questioned Documents.
12. Siegel S. Nonparametric statistics for the behavioral sciences. New York: McGraw Hill, 1956;122.
13. Sachs L. Applied statistics—a handbook of techniques. New York: Springer Verlag, 1984;§3.9.3.

Additional information and reprint requests:
Moshe Kam
ECE Department
Data Fusion Laboratory
3141 Chestnut Street
Philadelphia, PA 19014