Journal of the American Society of Questioned Document Examiners, Inc.
Volume 14, Number 1, June 2011

The Use of Principal Component Analysis to Provide Objective Methods for the Examination of Arabic Signatures

Ahmed Al Haddad[1] , Peter C. White[2] and Anthony M.Cowell[3]

[1] Abu Dhabi Police, PO Box 253, United Arab Emirates
[2] ReForensics, PO Box 1208, Lincoln, LN5 5LT, UK (Author for Correspondence).
[3] School of Natural and Applied Sciences, University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, UK.

The natural variation within an individual's Arabic signature is studied using Principal Component Analysis (PCA) with orthogonal rotation of the components. Fifty replicate analyses of sixteen features, measurable or classified by a numerical coding system, were analysed for a legible and an illegible type of signature. With either type of signature this PCA method provides an objective, robust and simple routine for identifying the features and importantly associations between features. Using only the measurable features, the results from two trials proves the method could potentially assist in confirming the authenticity and identifying forgeries for both legible and illegible types of signatures. From studying the effect of sample size it is recommended that no less than twenty-five control signatures should be used to ensure reproducibility and confidence in reporting any results when using this statistical method of analysis.

Introduction

In practical situations when a Forensic Document Examiner (FDE) has to determine if a signature is a forgery they will obtain multiple samples (control samples) of the genuine signature. One of the primary methods to compare the signatures is then to identify characteristic features in the genuine signature from the control samples and after taking into consideration the natural variation observed, determine if there are any significant differences between these features in the genuine and suspect signature. In selecting these features, FDEs are also searching for any associations between these features in the control samples as these will be characteristic of the genuine signature and a forger would have more difficulty in reproducing. The selection of the most discriminatory features and especially those where there may be associations is therefore very important and this has been recognised as a problem, even by experienced FDEs, throughout the history of document examination. It is also a very consuming and subjective process and furthermore, there have been concerns expressed in courts who favour the use of objective evidence. The desire to use more objective methods also appears to be supported by document examiners. For example, Blake 1977 has supported the move to employing objective measurements of stable handwriting forms and has concluded that "unless we attempt to validate our methods, however difficult that may be, we are not advancing as a profession".

There are many instances in forensic examinations where multivariate analytical methods have been used to provide objective evidence, especially in cases where proof is being sought to show samples have a originated from a common source. Examples include, drug profiling where typically drug, precursor and elemental compositions and concentration data are analysed or classification of glass samples based upon elemental composition and concentration data. Regarding the examination of documents, only very minimal use of multivariate analytical techniques has been reported. In one study discriminant analysis was used (Marquis and others 2005) to assess the variability in the shape of a handwritten capital letter and determine if it possible to discriminate between writers. Fourier descriptors, which provide a morphological characterization of closed contours (shape description), were evaluated by the examination of an individual handwritten capital character eg., the letter O". A total of 445 pieces of script containing this character were produced by three people. Pair-wise comparisons showed significant differences in the mean values of the Euclidean distances between each pair of writers at P< 0.001. Results from discriminant analyses showed no overlapping of the features between the samples produced by the three people. Hence this method may provide a means for discriminating between writers but it requires further testing with a much larger sample population.

As the aim of this study was to identify methods which could provide objective evidence in the examination of Arabic signatures, the use of Principal Component Analysis (PCA) was investigated since this multivariate analytical procedure can, from a set of data, be used to identify variables (features) and associations between variables (components). Additionally, PCA ranks the variables thus the technique should help to establish which ones produce the most variation between features. However, as it has been previously shown that natural variation in some features can be influenced by the legibility of the signature (Al Haddad and others 2009) and therefore, PCA analyses of both legible and illegible types of signatures were studied.

For the testing of the PCA method in this study fifty samples of an illegible signature from one person were used. To test if this PCA method could, in principle, work for determining the authenticity of a signature, a further three samples from this person were used. A more practical test of the PCA method was also undertaken to establish if it could be used to correctly identify known forgeries of this signature.

These studies were based upon the use of fifty control samples of the person's signature but in casework, the FDE typically uses about fifteen to twenty control samples. Hence the PCA studies were repeated using data from sets of twenty-five and sixteen drawn from the fifty samples used in the initial PCA study to determine the effect of sample size. Based upon these results thirty control samples of a legible signature were subjected to PCA and a further three were used to determine the authenticity of the signature. Finally, as with illegible signatures, PCA was used to determine if it could be used to correctly identify forgeries of legible signatures. The results of these studies are presented and discussed.

Method and Materials

Selection of Illegible Signature and Measurement of Features

The collection of Arabic signatures obtained for a previous study (Al Haddad and others 2009), were used again in this study. From within the collection of signatures one of the illegible type of signatures was selected randomly (using a random number generator) and fifty signatures produced by the participant were provided. One of these fifty samples of this signature was used to identify sixteen features that were selected for the PCA study. As shown in Figure 1, it can be observed that thirteen features can be measured directly from the signature. The other three are variables that are classified in a numerical coding system. Note all the measurements and classifications are based upon Arabic writing going from right to left.

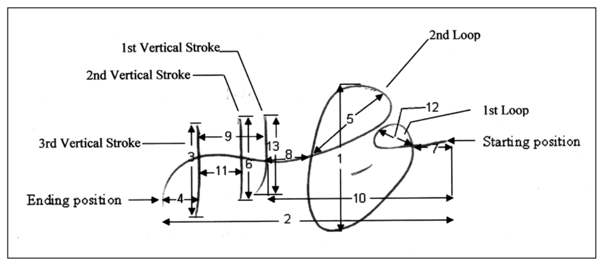

Figure 1 A sample of one of the fifty illegible signatures selected for the study showing the thirteen measured variables.

These thirteen features were measured using a digital Vernier caliper and identified as follows.

1. Height of signature; measured between the highest and lowest points of a signature.
2. Length of signature; measured from the beginning to the end of the signature.
3. Height of longest vertical stroke; based on length of the longest of the three vertical strokes in the signature.
4. Distance from end of signature to the 3rd vertical stroke.
5. 2nd loop diameter, measured at widest point.
6. Height of 2nd vertical stroke.
7. Distance of 1st loop from start of the signature.
8. Distance from 2nd loop to the 1st vertical stroke.
9. Distance between the 1st and 3rd vertical strokes.

10. Distance from start of signature to the 1st vertical stroke.
11. Distance between 2nd and 3rd vertical stroke.
12. 1st loop diameter, measured at widest point.
13. Height of 1st vertical stroke.

The other three features (classified variables) were calculated based on three scoring categories. These features were as follows:

14. Relationship between 1st and 2nd loop; Scored as follows:
    (Score 1) Open - where there is a space between the two loops.
    (Score 2) Closed - where there is no space between the two loops.
    (Score 3) Crossing - where the two loops are crossing each other.

Examples of signatures showing these different classifications are illustrated as shown in Figure 2.

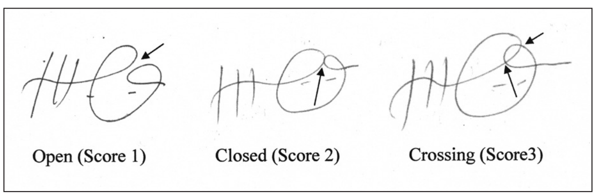

Figure 2 Examples of the illegible signature showing the different relationships between the 1st and 2nd loops and their scores.

15. Top of 1st vertical stroke in relation to the top of 2nd vertical stroke; Scored as follows:
    (Score 1) Below - where the top of the 1st vertical stroke is below the top of the 2nd vertical stroke.
    (Score 2) Similar - where the top of both 1st and 2nd vertical strokes are at the same level.
    (Score 3) Above - where the top of the1st vertical stroke is above the top of the 2nd vertical stroke.

Note that for this particular category, the vertical measurement was obtained as follows: Between the two vertical strokes there is a horizontal line and the point of intersection with the right vertical stoke with this line is taken as the reference for the baseline. The measurement of the vertical stroke is then the distance that the stroke extends vertically above this baseline. If the difference in the height of the two vertical strokes was ≤ 0.50mm, then these two vertical strokes were classified as being similar. Examples of signatures showing these different classifications are illustrated in Figure 3.

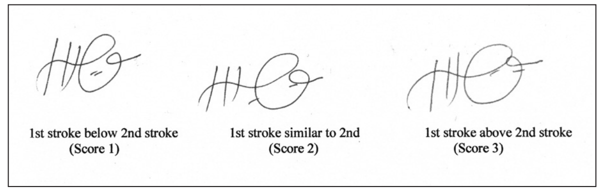

Figure 3 Examples of the illegible signature showing the different relationships between the heights of the 1st and 2nd vertical strokes, together with their scores.

16. Top of 3rd vertical stroke in relation to the top of the 1st vertical stroke; Scored as follows:
    (Score 1) Below - where the top of the 3rd vertical stroke is below the top of the 1st vertical stroke.
    (Score 2) Similar - where the top of 1st and 3rd vertical strokes are at the same level. This classification was determined in a manner similar to that described for feature 15.
    (Score 3) Above - where the 3rd vertical stroke is above the 1st vertical stroke.

Examples of signatures showing the different relationships between the top of the 1st and 3rd vertical strokes are shown in Figure 4.

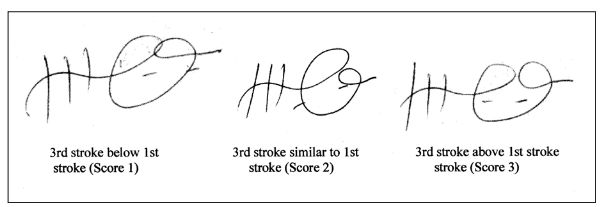

Figure 4 Examples of the illegible signature showing the different relationships between the heights of the 1st and 3rd vertical strokes, together with their scores.

Forged Illegible Signatures

To test if in principle the PCA method could be used for determining the authenticity of an illegible signature, three copies of this genuine signature, additional to the fifty used in the

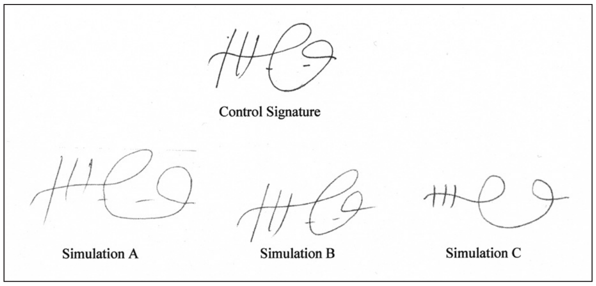

Figure 5 Copy of one of the control illegible signatures and the forged signature produced by three other writers.

initial PCA study were obtained from the participant. For the PCA study the same sixteen features in the genuine signature were identified and measured in these forged signatures.

To test the PCA method for determining the authenticity of a signature, a genuine signature from the person who was used to test the PCA was scanned at the original size and printed on plain paper. A copy of this was then given to three people (these people were not participants who provided signatures for the original collection) who were requested to provide one forgery of this signature. The copy of the genuine signature and the forged signature produced by each writer are shown in Figure 5.

Sub–Grouping of Signatures

To study the effect of sample size, sub-groups of the fifty signatures, numbered 1-50, were created. The sub-groups were the first twenty five signatures (1-25), the second twenty-five signatures (26-50), the even numbered and odd numbered signatures. A further sub-group of sixteen signatures was obtained by selecting every third signature ie., 1, 4, 7 etc.

Selection of Legible Signature and Measurement of Features

The collection of Arabic signatures obtained for the previous study (Al Haddad and others 2009), were used again in this study. From within the collection of signatures one of the legible type of signatures was selected randomly (using a random number generator) and thirty signatures produced by the participant were provided. As shown in Figure 6, it can be observed that sixteen features can be measured directly from the signature Note all the measurements are based upon Arabic writing going from right to left.

These sixteen features are identified as follows.

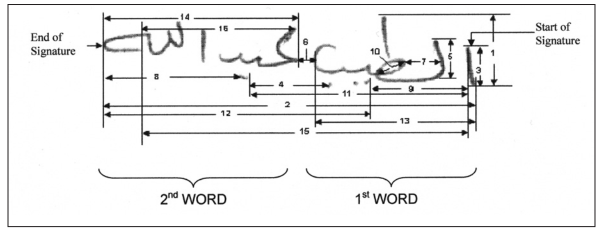

Figure 6 An example sample of one of the thirty legible signatures selected for the study showing the sixteen measured variables.

1. Height of signature.
2. Length of signature.
3. Height of 1st vertical stroke in the 1st word.
4. Distance between 3rd and 4th dot.
5. Height of 2nd vertical stroke in the 1st word.
6. Distance between the two words.
7. Distance between 2nd vertical stroke and loop in the 1st word.
8. Distance from end of the 4th dot to the end of the signature.
9. Distance from start of signature to 1st dot.
10. Loop diameter in the 1st word.
11. Distance from start of signature to 4th dot.
12. Distance between 1st dot and end of signature.
13. Length of 1st word.
14. Length of 2nd word.
15. Distance from start of the signature to the end of the last stroke in the signature.
16. Distance between start of 2nd word and the end of the last stroke in the signature.

Forged Legible Signatures

To test if in principle the PCA method could be used for determining the authenticity of a legible signature, three copies of this genuine signature, additional to the thirty used in the initial PCA study were obtained from the participant.

To test the PCA method in a more practical way, a genuine signature from the person who was used to test the PCA was scanned at the original size and printed on plain paper. A copy this was then given to three people (these were the same people who provided the forged illegible signatures in the previous study) who were requested to provide one forgery of this signature. The copy of the genuine signature and the forged signature produced by each writer are shown in Figure 7. For the PCA study the same sixteen features in the genuine signature were identified and measured in these forged signatures.

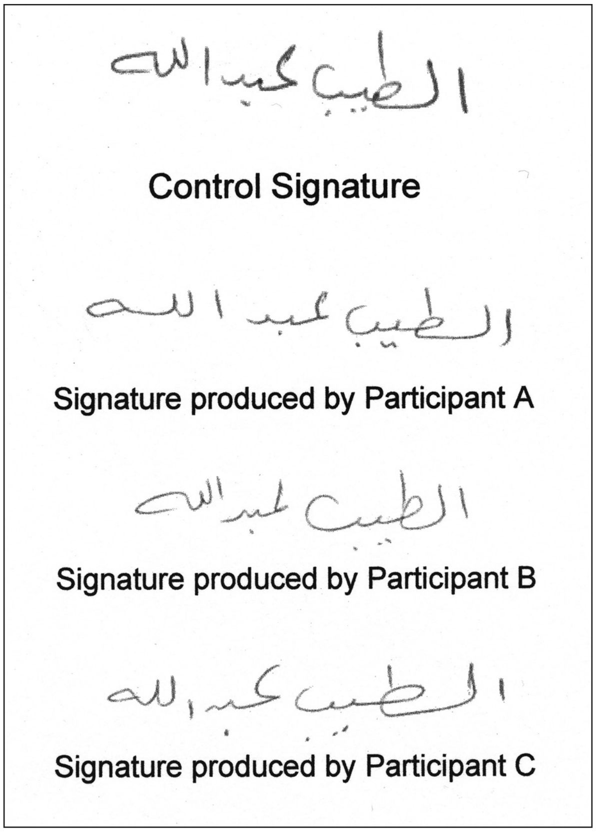

Figure 7 Copy of the genuine signature and the forged signature produced by each writer are shown below.

Statistical Analyses
All data collected for these studies were entered into a spreadsheet in SPSS (Statistical Package for the Social Sciences - Version 14, 2005) to aid the analysis. Checks for outliers and extreme scores were made using boxplots. To check for outlier effects, the differences between the means and the trimmed means for each variable were examined. Prior to carrying out PCA, the Kaiser-Meyer-Olkin (KMO) and Bartlett's test of Sphericity were also conducted to test for sampling adequacy.

PCA analyses were performed using the SPSS software to initially extract the components and then utilising the Varimax procedure to rotate the components identified in the PCA un-rotated correlation matrix.

Results and Discussion

PCA of an Illegible Signature
The sixteen features for all fifty replications of the illegible signature obtained in this study were measured and classified as described in the above Section, and these data were entered into the SPSS PCA statistical package. Initially, the descriptive statistics for each of the sixteen features for all the fifty samples of the signature were obtained (range, mean, standard deviation (SD) and coefficient of variation (%CV) to determine if variation for any of the features could be observed. The results obtained are shown in Table 1.

The visual examination of these data confirmed that for any of the features that had been selected, there was considerable variation, with a CV of 9.0 to 39.6 %. It is also interesting to note that the greatest variations were recorded with those features that were of the classified type of variable. This was expected since the classification scheme limits the number of forms to only three and these being recorded as integer values.

PCA was conducted to investigate if any relationships existed between the 16 features. However, prior to performing this analysis, the Kaiser-Meyer-Olkin (KMO) and Bartlett's test of Sphericity were used to test for the measuring of sampling adequacy. The KMO test produced a value of 0.633, and as it produced a P value of > 0.5, this test indicated from work published by Kaiser (1958) that the PCA would be an appropriate form of analysis and should yield distinct and reliable components,. The Bartlett's Test of Sphericity additionally confirmed statistical significance of the results with a P value of ≤ 0.000 being obtained.

PCA initially requires component extraction to find a linear combination of variables (a component) that accounts for as much variation in the original variables as possible for the illegible signature. As part of this process it is necessary to determine the number of extracted components and hence variables that could potentially be used. This was achieved by obtaining the Eigenvalues of the principal components and their absolute and relative magnitudes, and the results obtained are summarized in Table 2.

From these results it can be observed that there are as many components as there are features (variables) in the data set (i.e., 16 variables giving 16 components). Ideally, the first few components should account for at least 70-80% of the cumulative variance but in this analysis the first three components account for only 52% of the cumulative variance. The reason that three components accounted for only 52% of the cumulative variance was because the variables were not highly correlated. To identify the ideal number of components that should be retained,

NAMES OF VARIABLES		RANGE	MEAN	STD. DEVIATION	% CV
Measured Variables					
1.	Height of signature	15.06	23.13	3.20	13.83
2.	Length of signature	25.05	47.21	4.29	9.08
3.	Height of longest vertical stroke	11.88	19.27	2.64	13.70
4.	Distance from end of signature to the 3rd vertical stroke	11.14	5.93	1.99	33.55
5.	2nd loop diameter	9.53	17.01	2.49	14.63
6.	Height of 2nd vertical stroke	9.71	13.60	2.21	16.25
7.	Distance of 1st loop from start of the signature	7.82	6.43	1.74	27.06
8.	Distance from 2nd loop to the 1st vertical stroke	7.55	5.85	1.90	32.47
9.	Distance between the 1st and 3rd vertical strokes	7.11	9.10	1.61	17.69
10.	Distance from start of signature to the 1st vertical stroke	14.50	31.17	2.82	9.04
11.	Distance between 2nd and 3rd vertical stroke	4.46	4.95	0.96	19.39
12.	1st loop diameter	9.97	6.74	2.13	31.60
13.	Height of 1st vertical stroke	10.77	14.12	2.45	17.35
Classified Variables					
14.	Relationship between 1st and 2nd loop	2	2.24	0.85	37.94
15.	Top of 1st vertical stroke in relation to the top of 2nd vertical stroke	2	2.30	0.91	39.56
16.	Top of 3rd vertical stroke in relation to the top of 1st vertical stroke	2	2.30	0.86	37.39

Table 1 Descriptive statistics for each variable following the examination of fifty repetitions of the illegible signature. All data for the measured variables are stated in mm.

COMPONENT	EIGENVALUE	% OF VARIANCE	%CUMULATIVE VARIANCE
1	4.0	25.2	25.2
2	2.2	13.7	38.9
3	2.1	12.9	51.8
4	1.6	10.1	61.9
5	1.2	7.7	69.6
6	1.1	7.0	76.7
7	0.9	5.6	82.2
8	0.8	5.1	87.3
9	0.5	3.4	90.7
10	0.4	2.8	93.5
11	0.3	1.8	95.3
12	0.2	1.3	96.7
13	0.2	1.2	97.9
14	0.2	0.9	98.8
15	0.1	0.8	99.6
16	0.1	0.4	100.0

Table 2 Summary of components and their Eigenvalue data obtained from 50 replicate analyses of the sixteen features of an individual's illegible signature.

VARIABLE	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
11	0.77					
9	0.77					
2	0.72					
1	0.71					
5	0.69					
10	0.59					
3	0.55		-0.54			
6	0.53	-0.52				
7		0.73				
13		-0.63				
14		0.57				
8			0.65			
16			-0.50	0.67		
12				0.61		
15			0.52		0.70	
4			**0.42**			**-0.44**

Table 3 Results of PCA analysis showing the loading (correlation) between a variable and each of the six components. Negative values indicate an inverse correlation. Note except for Variable 4 (shown in bold), correlation values only greater than ± 0.5 are recorded.

the Eigenvalue Rule can be used and this stipulates that only components with an Eigenvalue of greater than 1.0 should be retained. Thus with this set of results, components 1-6 were retained.

The component matrix obtained to determine the loading (correlation) between a variable and each of its components is shown in Table 3. To simplify the results the method of only displaying loadings of greater than of > ± 0.5 was adopted as recommended by Chan (2004).

These results show clearly that half of the sixteen selected variables produced a loading value of greater than ± 0.5 and appear in Component 1, thus indicating correlation between these variables. Further examination of the other components show that fewer variables were correlated, either directly or inversely (negative values), and tend to decrease with an increase in the component number. It was also noted by applying a correlation cut-off value of > 0.5 that Variable 4 did not feature in any of the Components but on reducing the cut-off value to > 0.4, this variable was found in Component 3 (0.42) and 6 (-0.44).

From these results it was difficult to establish which variables are the most suitable for studying the signature of an individual due to the high degree of correlation observed between large numbers of the variables. Hence, a method for trying to improve this situation, namely by rotation of the components, was investigated.

The Varimax component rotation technique

COMPONENT	EIGENVALUE	% OF VARIANCE	%CUMULATIVE VARIANCE
1	2.8	17.8	17.8
2	2.3	14.6	32.4
3	1.9	11.7	44.0
4	1.8	11.5	55.6
5	1.8	11.4	67.0
6	1.6	9.7	76.7

Table 4 Eigenvalue data obtained after rotation of the first six components identified in the un-rotated PCA results.

was used to study the effect of rotation of the first six components obtained from the un-rotated PCA and initially their new Eigenvalues were computed. These, together with their variances are presented in Table 4.

Comparison of these results with those obtained for the un-rotated data (Table 2) show several interesting changes. Firstly, the Eigenvalues for Components 5 and 6 have increased significantly but the Eigenvalue for Component 1 has been reduced from 4.0 to 2.8. Rotation has also reduced the cumulative variance of each of the first five components but overall these six components still account for 76.7% of the cumulative variance. These results establish that the six components should be used for the rotation analysis and more clearly defined groups of variables for each component would be expected and this was confirmed as shown in Table 5. Again only variables with a correlation value of > ± 0.5 are reported.

The results presented in Table 5 when compared with those obtained for the un-rotated analysis (Table 3), show that rotation has substantially reduced the number of correlated variables for each component and in particular for Component 1. In nearly all instances rotation has also produced higher correlation values and when compared with the un-rotated results, the number of variables loaded into more than one component has been reduced. The only incidence where this occurred was with Variable 1 which can be seen in both Components 1 and 5. However, as reported, this variable can be excluded from Component 5 because it was displayed with a higher correlation value in Component 1. Furthermore, it can be observed that Variable 4 is now featured and displays a correlation value of -0.61 and could have been excluded since in the un-rotated PCA, it had a correlation value below the acceptance level of ±0.05.

Clearly, rotation of the components has created

VARIABLE	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
9	0.88					
11	0.83					
1	0.63				0.53	
5	0.59					
10		0.87				
2		0.80				
7			0.83			
14			0.67			
13				0.84		
6				0.74		
4				-0.61		
8					-0.74	
3					0.71	
12					0.55	
15						0.87
16						-0.79

Table 5 Results of PCA analysis of data from an illegible signature showing the loading (correlation) between a variable and each of the six components after rotation. Negative values indicate an inverse correlation and correlation values only greater than ± 0.5 were recorded.

a more simplified matrix which should help to provide an insight into those features that can possibly identify variation within the signature of an individual. Furthermore, and of considerable benefit to the FDE is the fact that PCA can provide an objective method for identifying those variables that are associated (eg., having the same component number) and the ranking of associated features (eg., their component number) as illustrated in Table 6.

Having reached this stage, it is usual practice when using PCA to name each component to assist in the interpretation of the results. This is usually achieved by trying to find a name that will adequately describe all the correlated variables related to a specific component. In practice the FDE would be using this PCA method to determine if a signature was genuine or a forgery. Since any individual's signature will have its own unique set of variables which will not be the same as for this signature currently being studied, this particular set of data will not be re-used. Thus the standard procedure for naming components is thus unnecessary and therefore not recommended.

Examination of Forgeries of Illegible Type of Signature

To confirm that PCA could help in determining if the procedure could be use in the process of identifying if signature is a forgery two studies

COMPONENT No.	VARIABLE	
	No.	DESCRIPTION
1	9	Distance between the 1st and 3rd vertical strokes
	11	Distance between 2nd and 3rd vertical stroke
	1	Height of signature
	5	2nd loop diameter
2	10	Distance from start of signature to the 1st vertical stroke
	2	Length of signature
3	7	Distance of 1st loop from start of the signature
	14	Relationship between 1st and 2nd loop
4	13	Height of 1st vertical stroke
	6	Height of 2nd vertical stroke
	4	Distance from end of signature to the 3rd vertical stroke
5	8	Distance from 2nd loop to the 1st vertical stroke
	3	Height of longest vertical stroke
	12	1st loop diameter
6	15	Top of 1st vertical stroke in relation to the top of 2nd vertical stroke
	16	Top of 3rd vertical stroke in relation to the top of 1st vertical stroke

Table 6 PCA results for the illegible signature which identify the descriptions of the variables that are associated (same component number) and their ranking.

were undertaken. Firstly, a further three test signatures (A,B and C) from the person used in the above PCA study were used to illustrate how PCA could be used for detection of a forgery. The second test was to take signatures known to be forged and hence examine the potential of this method.

To be able to use this PCA method in a casework scenario, ie., to determine the authenticity of a suspect signature, an FDE would require the PCA results and the maximum and minimum values for each variable from a set of genuine signatures, summarized as shown in Table 7. The results presented relate to the set of fifty analyses of the genuine illegible signature used in this particular study.

COMPONENT No.	VARIABLE		RANGE OF VARIABLE		TEST SIGNATURES		
	No.	DESCRIPTION	MINIMUM	MAXIMUM	A	B	C
1	9	Distance between the 1st and 3rd vertical strokes	5.52	12.63	8.38	9.37	10.91
	11	Distance between 2nd and 3rd vertical stroke	3.36	7.82	4.65	5.46	5.51
	1	Height of signature	17.48	32.54	17.81	22.47	22.69
	5	2nd loop diameter	12.50	22.03	14.99	15.57	15.92
2	10	Distance from start of signature to the 1st vertical stroke	23.35	37.85	28.34	33.95	34.12
	2	Length of signature	34.56	59.61	43.37	46.59	50.04
3	7	Distance of 1st loop from start of the signature	2.02	9.84	5.77	6.47	8.01
	14	Relationship between 1st and 2nd loop	1	3	3	3	2
4	13	Height of 1st vertical stroke	9.04	19.81	11.72	12.04	12.18
	6	Height of 2nd vertical stroke	8.89	18.60	12.03	12.26	14.21
	4	Distance from end of signature to the 3rd vertical stroke	1.78	11.14	2.38	4.51	5.71
5	8	Distance from 2nd loop to the 1st vertical stroke	1.94	9.49	3.73	5.28	7.75
	3	Height of longest vertical stroke	14.60	26.48	15.41	17.87	17.93
	12	1st loop diameter	2.38	12.35	6.46	6.66	10.66
6	15	Top of 1st vertical stroke in relation to the top of 2nd vertical stroke	1	3	3	3	3
	16	Top of 3rd vertical stroke in relation to the top of the 1st vertical stroke	1	3	1	1	3

Table 7 Summary of PCA results and the range of measurements for each variable obtained from the fifty genuine signatures used in this study, together with the measurements obtained from an additional three test signatures A, B and C produced by the original writer.

To determine the authenticity of a suspect signature is then a fairly simple task that initially requires measurements of the same sixteen features in the test signatures as for the genuine signatures. The measurement of each variable can then be compared with the data obtained for the control sigantures as demonstrated in Table 7. If the suspect signature is genuine then for any variable the measurement obtained should fall within the range of the measurement of the genuine sample.

The visual examination of these data confirmed that for any of the variables that had been measured for the three test signatures, each measurement fell within the range of the appropriate variable for the genuine samples of the fifty illegible signatures. Hence, these test signatures must have been produced by the same person, which is correct, as the test signatures were genuine signatures. These results therefore confirm that the procedure proposed for using PCA in the determining the authenticity of a signature does work.

To determine the potential of using PCA to identify a forged illegible signature, the sixteen features of the genuine signature were measured in the forged copy of the signature provided by three people (A, B and C). The measurements for each feature were obtained, and these were then compared with the data of the genuine signatures

COMPONENT No.	VARIABLE		RANGE OF VARIABLE OF CONTROL SIGNATURES		FORGED SIGNATURES		
	No.	DESCRIPTION	MINIMUM	MAXIMUM	A	B	C
1	9	Distance between the 1st and 3rd vertical strokes	5.52	12.63	14.85	11.35	3.66
	11	Distance between 2nd and 3rd vertical stroke	3.36	7.82	6.07	6.23	4.31
	1	Height of signature	17.48	32.54	33.33	29.86	18.34
	5	2nd loop diameter	12.50	22.03	23.58	20.59	11.66
2	10	Distance from start of signature to the 1st vertical stroke	23.35	37.85	44.13	35.49	43.34
	2	Length of signature	34.56	59.61	67.28	55.81	54.67
3	7	Distance of 1st loop from start of the signature	2.02	9.84	1.89	7.19	5.72
	14	Relationship between 1st and 2nd loop	1	3	1	1	1
4	13	Height of 1st vertical stroke	9.04	19.81	15.91	18.52	8.53
	6	Height of 2nd vertical stroke	8.89	18.60	18.41	16.97	7.65
	4	Distance from end of signature to the 3rd vertical stroke	1.78	11.14	7.99	5.44	3.43
5	8	Distance from 2nd loop to the 1st vertical stroke	1.94	9.49	7.08	11.62	9.21
	3	Height of longest vertical stroke	14.60	26.48	29.67	24.01	8.75
	12	1st loop diameter	2.38	12.35	15.70	10.28	9.27
6	15	Top of 1st vertical stroke in relation to the top of 2nd vertical stroke	1	3	3	1	3
	16	Top of 3rd vertical stroke in relation to the top of the 1st vertical stroke	1	3	3	1	1

Table 8 Comparison of the results between the genuine set of fifty signatures (Control) and the three forged signatures of three participants, Signatures A, B and C. The highlighted data indicates features where the value of a forgedsignature falls outside of the range of the genuine signature.

as in the model described above and analysed by PCA. The results obtained are presented in Table 8.

These results show that the three classified variables for test signatures all fall within the range of the genuine signature and therefore incorrectly identify these as possibly being genuine signatures. Hence, although these features illustrate variation within signatures, the coding system does not provide a method for discrimination between signatures in this PCA method and confirm that classified variables should not be used.

Examination of the measurable features show that with at least 6 of the 13 measurable variables for signatures A and C these fall out side of the ranges obtained from the genuine signature, and hence a strong indication that these signatures are forgeries. The results for signature B do not provide the same strength evidence of a forgery since only one variable falls outside of the range observed for the genuine signature. However, this result should certainly alert the FDE of a potential forgery and other factors such as fluidity of a signature should be used by the FDE to make a final decision. These results highlight that the success of PCA in this task will be dependent on the FDE's original choice of measurable features.

VARIABLE	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
9	0.77					
11	0.81					
1	0.71					
5	0.67					
10		0.90				
2		0.88				
7			0.83			
14			0.64			
13				0.91		
6				0.66		
4				-0.60		
8					-0.66	
3					0.59	
12					0.70	
15						0.88
16						-0.81

Table 9 Results for the first twenty five (1-25) signatures following PCA and component rotation.

With this particular signature, the use of other measurable features may also provide a higher level of discrimination and therefore, the FDE should not limit the number of features measured as PCA would help to identify the most suitable features for the comparison of signatures.

Effect of Sample Size

In the examination of signatures, the FDE would have typically a set of fifteen repeats of a signature (control samples) to select the features they would use for the comparison of signatures. Since this PCA method has a recommended sample size of fifty signatures, PCA analyses were repeated using smaller subsets of the same set of signatures to identify any effects arising from using smaller sample sizes.

To investigate the effect of sample size, subgroups of the fifty signatures, numbered 1-50, were created. The sub-groups were the first twenty-five signatures (1-25), the second twenty-five signatures (26-50), the even numbered and odd numbered signatures. A further smaller sub-group of sixteen signatures was obtained by selecting every third signature ie., 1, 4, 7 etc. Component loading patterns for these groups of samples used the PCA and rotation procedures outlined previously for the set of fifty signatures. The loadings patterns from the groups with a sample size of twenty five are shown in Tables 9 - 12.

The results for the smaller sample size of twenty-five were compared with data generated from the original set of fifty signatures (Table 7)

VARIABLE	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
(9)	0.92					
(11)	0.84					
(1)	0.77					
(5)	0.66					
(10)		0.58				
(2)		0.66				
(7)				0.58		
(14)				0.82		
(13)			0.84			
(6)			0.87			
(4)			-0.66			
(8)					-0.63	
(3)					0.87	
(12)						0.54
(15)						0.86
(16)						-0.80

Table 10 Results for the second twenty five (26-50) signatures following PCA and component rotation.

VARIABLE	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
(9)	0.87					
(11)	0.83					
(1)	0.74					
(5)	0.72					
(10)		0.55				
(2)		0.79				
(7)					0.76	
(14)					0.68	
(13)			0.86			
(6)			0.85			
(4)			-0.64			
(8)				-0.78		
(3)				0.74		
(12)				0.67		
(15)						0.90
(16)				-0.78		

Table 11 Results for the odd numbered group of twenty five signatures following PCA and rotation of the components.

and showed that all the correlation values were still greater than ± 0.5, with the Variables 4, 8 and 16 still retaining their inverse correlation. A noticeable feature was that Components 1 and 2 always retained the same variables in every sample set of signatures with no interchange of the groups of variables between components. However, apart from the 1-25 set of signatures interchanges of the same groups of variables between components can be observed in the other three sub-groups of signatures.

The only other changes observed in using the smaller sample size of twenty-five was that in two instances one variable has moved into another component ie., Variable 12 in the 26-50 set of signatures and Variable 16 in the odd numbered set of signatures. In both of these cases the overall effect would be considered to be quite minor as they are low ranking variables. When reducing the sample size further ie., down to sixteen samples, some minor additional variations in the loading patterns were observed, as illustrated in Table 13.

It can be observed that there are no changes with Components 1 and 2, but again there are interchanges of groups of variables between components. However, in this sample size one of the variables (Variable 8) has not retained its inverse correlation and has separated from the original group of variables ie., 8, 3, 12.

The original choice of using a sample size of fifty for PCA was based upon guidelines reported by Barrett and Kline (1981) and the results obtained using a smaller sample size of twenty-five for studying variations in signatures indicate only minor variations in the correlation matrix. By reducing the sample size down to sixteen, typical of that currently used by FDEs, some further minor changes were observed. These results suggest that a minimum sample size of twenty-five should be used for this type of analysis to ensure a high degree of confidence in reporting any result.

PCA Analysis of a Legible Signature

Based upon the PCA results obtained for the illegible signature no classified features were used in this study. Again based upon a recommendation from this study that the minimum number of control samples should not be less than twenty-five, thirty control samples of the genuine signature were used to identify and determine by PCA the variability of the sixteen measurable features identified in this control signature.

The sixteen features for all thirty replications of the legible signature obtained in this study were measured as outlined in Methods and Materials and these data were entered into the SPSS PCA statistical package. The descriptive statistics for each of the sixteen features for all the thirty samples of the signature were obtained (range, mean, SD and %CV) to determine if variation for any of the features could be observed. The results obtained are shown in Table 14.

The visual examination of these data confirmed that for any of the features that had been selected, there was considerable variation with the CV

VARIABLE	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
(9)	0.90					
(11)	0.86					
(1)	0.67					
(5)	0.63					
(10)		0.63				
(2)		0.71				
(7)						0.65
(14)						0.91
(13)				0.87		
(6)				0.79		
(4)				-0.56		
(8)					-0.73	
(3)					0.83	
(12)					0.79	
(15)			0.90			
(16)			-0.78			

Table 12 Results for the even group of twenty five signatures following PCA and component rotation.

VARIABLES	LOADING FOR EACH COMPONENT					
	1	2	3	4	5	6
9	0.79					
11	0.59		0.54			
1	0.74					
5	0.88					
10		0.89				
2		0.89				
7					0.69	
14					0.74	
13			0.66			
6			0.92			
4			-0.64			
8					0.89	
3				0.86		
12				0.74		
15						0.92
16						-0.66

Table 13 Results for 16 signatures (every third signature) following PCA and rotation of the components.

	NAMES OF VARIABLES	RANGE	MEAN	STD. DEVIATION	%CV
1.	Height of signature	7.06	12.97	1.61	12.41
2.	Length of signature	13.76	38.08	4.03	10.58
3.	Height of 1st vertical stroke in the 1st word	3.51	6.07	0.923	15.20
4.	Distance between 3rd and 4th dot	6.18	3.82	1.15	30.10
5.	Height of 2nd vertical stroke in the 1st word	2.43	5.62	0.64	11.38
6.	Distance between the two words	3.66	1.62	0.79	48.76
7.	Distance between 2nd vertical stroke and loop in the 1st word	4.56	7.44	1.16	15.59
8.	Distance from end of the 4th dot to the end of the signature	11.58	12.53	2.94	23.46
9.	Distance from start of signature to 1st dot	9.46	16.18	2.33	14.40
10.	Loop diameter in the 1st word	4.42	6.32	1.02	16.13
11.	Distance from start of signature to 4th dot	7.82	23.74	2.27	9.56
12.	Distance between 1st dot and end of signature	11.87	21.79	3.00	13.76
13.	Length of 1st word	8.24	20.28	2.13	10.50
14.	Length of 2nd word	12.01	16.33	3.26	19.96
15.	Distance from start of the signature to the end of the last stroke in the signature	15.50	23.88	4.50	18.84
16.	Distance between start of 2nd word and the end of the last stroke in the signature	13.08	5.00	2.69	53.8

Table 14 Descriptive statistics for each variable following the examination of thirty repetitions of the legible signature. All data are stated in mm.

ranging 9.56 (Variable 11) to 53.8 (Variable 16). These results confirm that there is considerable variation in each of the selected features.

PCA was conducted to investigate if any relationships existed between the 16 features. Prior to performing this analysis, the Kaiser-Meyer-Olkin (KMO) and Bartlett's test of Sphericity were used to test for the measuring of sampling adequacy. The KMO test produced a value of 0.587, and as it produced a P value of > 0.5, this test indicated that the PCA would be an appropriate form of analysis and should yield distinct and reliable components. The Bartlett's Test of Sphericity additionally confirmed statistical significance of the results with a P value of ≤ 0.000 being obtained.

As described earlier, the PCA initially requires component extraction to find a linear combination of variables (a component) that account for as much variation in the original variables as possible for the legible signature. As part of this process it is necessary to determine the number of extracted components and hence variables that could potentially be used. This was achieved by obtaining the Eigenvalues of the principal components and their absolute and relative magnitudes and the results obtained are summarized in Table 15**.**

From these results it can be observed that there are as many components as there are features (variables) in the data set (i.e., 16 variables giving 16 components). Ideally, the first few components should account for at least 70-80% of the cumulative variance but in this analysis the first three components account for only 65% of the cumulative variance. The reason that the

COMPONENT	EIGENVALUE	% OF VARIANCE	CUMULATIVE % OF VARIANCE
1	5.3	33.0	33.0
2	3.3	20.5	53.4
3	1.9	11.6	65.1
4	1.5	9.1	74.2
5	1.2	7.7	81.8
6	0.9	5.7	87.6
7	0.7	4.4	92.0
8	0.5	2.9	94.9
9	0.4	2.4	97.2
10	0.2	1.2	98.5
11	0.1	0.8	99.3
12	0.1	0.3	99.7
13	0.0	0.2	99.8
14	0.0	0.1	99.9
15	0.0	0.1	100.0
16	0.0	0.0	100.0

Table 15 Summary of components and their Eigenvalue data obtained from 30 replicate analyses of the sixteen features of an individual's legible signature.

VARIABLES	LOADING FOR EACH COMPONENT				
	1	2	3	4	5
2	0.91				
11	0.86				
9	0.85				
13	0.84				
3	0.68				
12	0.55	**0.79**			
14	0.55	**0.78**			
8	0.51	**0.76**			
5		-0.58			
4			0.85		
6			0.81		
16				0.75	
10	*0.46*			***0.48***	
7	*0.46*		***-0.40***	***-0.47***	
1	*0.43*				**0.72**
15	0.53				**-0.55**

Table 16 Results of PCA analysis showing the loading (correlation) between a variable and each of the five components. Negative values indicate an inverse correlation. Note correlation values greater than ± 0.4 are shown in italics and the bold figures indicate variables that occur in another component.

first three components account for only 65% of the cumulative variance is because the variables were not highly correlated.

To identify the ideal number of components that should be used, the Eigenvalue Rule was applied with Components 1-5 being retained. A Component matrix was produced to show the correlation between a variable and each of the five Components and this can be seen in Table 16. From examination of these results it can be observed twelve of the sixteen variables appear in Component 1 and nine of these produced a loading value of greater than ± 0.5, thus indicating a high degree of correlation between these variables. Other components show that fewer variables were correlated, either directly or inversely (negative values), and tend to decrease with an increase in the component number. Additionally, as the highlighted values indicate, several of the features correlate across multiple components.

Having shown the benefits of carrying out a Varimax rotation with the illegible signatures these data for the legible signature were taken through the same rotation procedure and the Eigenvalue data obtained are shown in Table 17.

Comparison of these results with those obtained for the un-rotated data (Table 15) show several interesting changes. Firstly, the Eigenvalues for Components 3, 4, and 5 have increased significantly but the Eigenvalue for Component 1 has been reduced from 5.3 to 3.9 and rotation has also produced a similar variance for Component 2. These results suggest that these five components should all be retained and that the rotated PCA data should produce more easily identifiable groupings of variables. The Component matrix showing the correlation between the variables and the components following rotation are shown in Table 18.

The results presented in Table 18 when compared with those obtained for the un-rotated analysis (Table 16) show that rotation has substantially reduced the number of correlated variables for each component and in particular for Component 1. One of these variables appears with a higher correlation value in Component 2 and as reported previously it can be omitted from Component 1. Two further differences can also be observed. Firstly, three of the variables were inversely correlated but on rotation all variables are directly correlated. Secondly, with the un-rotated results it had been shown that Variables 7 and 10 were not loaded into any component unless the correlation cut-off value was reduced from ± 0.5 to ± 0.4. However by including these variables, rotation has considerably increased their correlation values with Variable 7 loading highly into Component 1 and Variable 10 loading

COMPONENT	EIGENVALUE	% OF VARIANCE	%CUMULATIVE VARIANCE
1	3.9	24.4	24.4
2	3.8	23.4	47.8
3	2.2	13.5	61.4
4	1.9	11.8	73.2
5	1.4	8.7	81.8

Table 17 Eigenvalue data obtained after rotation of the five components identified in the un-rotated PCA results.

VARIABLES	LOADING FOR EACH COMPONENT				
	1	2	3	4	5
9	0.89				
7	0.81				
13	0.80				
11	0.77				
3	0.64				
12		0.97			
8		0.96			
14		0.96			
2	0.51	0.81			
12			0.87		
10			0.67		
5			**0.43**		
1				0.88	
4				0.86	
1					0.79
16					0.78

Table 18 Results of PCA analysis showing the loading (correlation) between a variable and each of the five components after rotation. Except for Variable 5 shown in bold, correlation values of only greater than ± 0.5 are recorded.

into Component 3. Hence if rotation had not been used, these variables may have been excluded which would lead to a loss of discrimination between signatures.

Clearly PCA analysis and rotation of the components has created a more simplified matrix and this can be observed when the results are tabulated as shown in Table 19. These results also show those features that give rise to variation within the signature of an individual and the associations between the features. For reasons described in the previous article no attempt was made to name the components.

Based upon practical experience in comparing these signatures, the selection of features and identifying which features are associated is complex, difficult and time consuming and this is a subjective process. With this legible signature which is more complex than the illegible one studied previously, this will take the FDE much more time to identify features and present many more difficulties in trying to identify which features may be associated. The use of PCA to complete this task has therefore considerably reduced these problems and furthermore, if comparing signatures, it is an objective method and the results provide a ranking of associated features (components) identifying those that could be used to show the level of similarity/dissimilarity between two signatures.

Examination of Forgeries of Legible Type of Signature

To confirm that PCA could help in determining if the procedure could be used in the process of determining the authenticity of a signature two studies similar to those described for the illegible signature were undertaken.

In the first of these studies three test signatures (A, B and C) from the person who provided the legible signature were used to illustrate how PCA could be used for detection of a forgery. As previously described this requires the PCA results and the maximum and minimum values for each variable from a set of genuine signatures, and these are summarized as shown in Table 20. The results presented relate to the set of thirty analyses of the genuine legible signature used in this particular study.

As described previously, to determine the authenticity of a suspect signature this is a fairly simple task that initially requires measurements of the same sixteen features for the test signatures. The measurements for each variable for the test signatures can then be compared with the data from the genuine signatures as demonstrated in Table 20.

The visual examination of these data confirmed that for any of the variables that had been measured for the three test signatures, each measurement fell within the range of the appropriate variable for the genuine samples of the thirty legible signatures. Hence, these test signatures must have been produced by the same person, which is correct, as the test signatures were genuine signatures. These results therefore confirm the potential of using this PCA based procedure to determine the authenticity of a signature.

The testing of the PCA procedure for detecting a forged legible signature was carried out as described for the illegible type of signature. The sixteen features of the genuine signature were

COMPONENT No.	VARIABLE	
	No.	DESCRIPTION
1	9	Distance from start of signature to 1^{st} dot
	7	Distance between 2^{nd} vertical stroke and loop in the 1^{st} word
	13	Length of 1^{st} word
	11	Distance from start of signature to 4^{th} dot
	3	Height of 1^{st} vertical stroke in the 1^{st} word
2	12	Distance between 1^{st} dot and end of signature
	8	Distance from end of the 4^{th} dot to the end of the signature
	14	Length of 2^{nd} word
	2	Length of signature
3	15	Distance from start of the signature to the end of the last stroke in the signature
	10	Loop diameter in the 1st word
	5	Height of 2^{nd} vertical stroke in the 1^{st} word
4	6	Distance between the two words
	4	Distance between 3^{rd} and 4^{th} dot
5	1	Height of signature
	16	Distance between start of 2^{nd} word and the end of the last stroke in the signature

Table 19 Summary of PCA results for the legible signature which can be used to identify the features that show most variation and associations between features.

measured in the forged copy of the signature provided by three people (A, B and C) and the measurements for each feature were obtained. These were then compared with the data of the genuine signatures as in the model described above and analysed by PCA. The results obtained are presented in Table 21.

These results clearly indicate that the three signatures are forged copies of the original signature thus showing the potential of using this PCA based method. This conclusion is based upon several observations. Firstly, the signatures all have seven or eight features that appear in the first three components for the genuine signature and these three components accounted for 65% of the total variance of all the features. Secondly, each of the signatures show some feature outside of the range of the genuine signature for the other lower ranked components. Finally, with the genuine signature when tested against the control samples, none of the features were outside of the range of natural variation of the genuine signature for any of the features.

COMPONENT No.	VARIABLE		RANGE OF VARIABLE		TEST SIGNATURES		
	No.	DESCRIPTION	MINIMUM	MAXIMUM	A	B	C
1	9	Distance from start of signature to 1^{st} dot	11.19	20.65	14.04	14.56	16.62
	7	Distance between 2^{nd} vertical stroke and loop in the 1^{st} word	5.26	9.82	5.68	5.59	8.86
	13	Length of 1^{st} word	15.99	24.23	20.16	19.51	21.93
	11	Distance from start of signature to 4^{th} dot	19.48	27.30	21.02	22.68	23.36
	3	Height of 1^{st} vertical stroke in the 1^{st} word	4.73	8.24	6.78	6.30	7.12
2	12	Distance between 1^{st} dot and end of signature	16.08	27.95	21.74	23.60	21.81
	8	Distance from end of the 4^{th} dot to the end of the signature	7.73	19.31	13.26	13.47	13.02
	14	Length of 2^{nd} word	10.44	22.45	17.16	17.94	17.07
	2	Length of signature	32.19	45.95	38.98	40.21	40.11
3	15	Distance from start of the signature to the end of the last stroke in the signature	15.54	31.04	26.74	24.61	26.27
	10	Loop diameter in the 1st word	4.16	8.58	6.63	6.80	5.71
	5	Height of 2^{nd} vertical stroke in the 1^{st} word	4.44	6.87	5.87	5.30	4.88
4	6	Distance between the two words	0.80	4.46	1.45	1.09	1.22
	4	Distance between 3^{rd} and 4^{th} dot	1.60	7.78	4.76	5.05	4.51
5	1	Height of signature	10.37	17.43	12.41	14.90	12.54
	16	Distance between start of 2^{nd} word and the end of the last stroke in the signature	1.37	14.45	5.50	3.08	5.53

Table 20 Summary of PCA results and the range of measurements for each variable obtained from the thirty genuine signatures used in this study, together with the measurements obtained from an additional three test signatures A,B and C produced by the original writer.

When compared with the results obtained for the illegible signature, it appears that the PCA method has been more successful with the legible signature. Two possible reasons are that the legible signature was more complex and hence more difficult to copy and/or that a larger number of features were measured in the legible signature. Overall it has been observed that legible signatures are more complex and thus more difficult to copy but further PCA studies would be required to confirm this observation and the effects of increasing the number of measurable features.

As indicated previously, examination of the features in a signature is only one of the methods used to establish the authenticity of a signature and other factors such as fluidity of a signature would be used by the FDE.

COMPONENT No.	VARIABLE		RANGE OF VARIABLE OF CONTROL SIGNATURES		SIMULATED SIGNATURES		
	No.	DESCRIPTION	MINIMUM	MAXIMUM	A	B	C
1	9	Distance from start of signature to 1st dot	11.19	20.65	18.52	11.32	24.78
	7	Distance between 2nd vertical stroke and loop in the 1st word	5.26	9.82	7.56	3.40	8.41
	13	Length of 1st word	15.99	24.23	27.70	26.98	34.55
	11	Distance from start of signature to 4th dot	19.48	27.30	38.14	35.50	44.30
	3	Height of 1st vertical stroke in the 1st word	4.73	8.24	6.61	5.66	3.97
2	12	Distance between 1st dot and end of signature	16.08	27.95	51.96	42.99	40.27
	8	Distance from end of the 4th dot to the end of the signature	7.73	19.31	32.19	19.13	19.23
	14	Length of 2nd word	10.44	22.45	40.42	25.32	28.28
	2	Length of signature	32.19	45.95	70.96	55.48	65.85
3	15	Distance from start of the signature to the end of the last stroke in the signature	15.54	31.04	63.09	46.99	59.60
	10	Loop diameter in the 1st word	4.16	8.58	4.30	4. 33	7.48
	5	Height of 2nd vertical stroke in the 1st word	4.44	6.87	5.24	5.46	5.95
4	6	Distance between the two words	0.80	4.46	3.16	2.99	2.46
	4	Distance between 3rd and 4th dot	1.60	7.78	7.65	12.31	7.75
5	1	Height of signature	10.37	17.43	10.51	10.78	11.35
	16	Distance between start of 2nd word and the end of the last stroke in the signature	1.37	14.45	31.21	14.40	21.95

Table 21 Comparison of the results between the genuine set of thirty signatures (Controls) and the three forged signatures of three participants, Signatures A, B and C. The highlighted data indicates features where the value of a forged signature falls outside of the range of the genuine signature.

However, these results highlight how PCA can be used to identify a high number of the most suitable features and those that are associated and this would be extremely difficult and time consuming if carried out manually for comparing legible signatures. As with the illegible type of signature, the degree of confidence in assessing if a signature is a forgery may be improved by increasing the number of measurable variables. With this particular signature, there were other measurable features that could have been used. Hence the FDE should not limit the number of features they select as PCA would identify features that are the most suitable for the comparison of signatures.

Conclusions

This study set out to determine if PCA offers the FDE an objective, fast and reliable method for exploring and identifying features that are variable in a signature and also associations between features. Through the use of PCA and by applying an orthogonal rotation method, it was shown that from an original selection of 16 features from replicate analyses of either an illegible or legible signature, PCA produced six clear components which accounted for 76.7% of the total variance and included all the features for the illegible signature. Similar results were obtained with the legible signature where 5 components accounted for 81.8% the total variance and included all the features for the legible signature. It was also shown that the selection of the appropriate cut-off level is an important consideration.

To study the effect of sample size it was found that, if using any one of the four subsets of twenty-five signatures, only some very minor changes in the component listings were observed. A further reduction to sets of 16 signatures produced some small changes. Hence, this study suggests that if considering the use of this PCA method for signature authentication, it is recommended that a minimum of twenty-five control samples of the signature are obtained to ensure the highest level of confidence in reporting results.

Further studies showed success was achieved in using PCA to confirm the authenticity of both the legible and illegible types of signatures, using samples that were both genuine and forged. It was established that only the measurable features can be used and the results indicated that PCA was more successful with the legible type of signature. However, differences in either the complexity of these signatures or the number of features measured could explain this observation. It was believed that further improvements in using PCA for establishing the authenticity of a signature may be gained by increasing the number of features that can be measured and further studies would confirm if either the complexity and/or number of features are controlling factors. The identification of features and their associations is only one procedure that an FDE uses to confirm the authenticity of a signature but PCA does remove the problem of subjectivity.

These preliminary studies indicate the potential for using PCA to examine signatures since it provides an objective method for identifying features, association of any of these features and their ranking of importance. Furthermore, it is much less time-consuming than the methods currently being used. These studies were however completed by using a limited number of signatures and clearly further testing is required to confirm the accuracy and robustness of this new PCA method for the examination of signatures. Workshops are currently being set up to undertake this work and the results obtained will be evaluated and presented in a future publication.

References

Al Haddad, A,; White, P.C. and Cole, M.D. (2009). Examination of a Collection of Arabic Signatures. *Journal of the American Society of Questioned Document Examiners.* Vol. 12, No.1, pp 35-53.

Barrett, P.T. and Kline, P. (1981). *Personality Study and Group Behaviour.* Vol 1, pp 23-33.

Blake, M. (1977) Handwriting Individuality: Can We Prove It? Presented at the American Society of Questioned Document Examiners Annual Meeting,

Chan, Y.H. (2004). *Biostatistics 302.* Vol. 4, No. 12, pp 558-565.

Kaiser, H. (1974). *Psychometrika.* Vol. 39, pp 31-36.

Marquis, R,; Schmittbuhi, M,; Mazzella, W.D. and Taroni, F. (2005). *Forensic Science International.* Vol. 150, pp 23-32.

Acknowledgements

The authors would like to thank the support of the Abu Dhabi Police and Professor Michael Cole (Anglia Ruskin University), John Flynn and Dr. Dorothy Gennard (University of Lincoln) and Tony Stockton (Forensic Science Service) for their advice and time spent reviewing our statistical data.