
**Pacific Northwest
National Laboratory**

Operated by Battelle for the
U.S. Department of Energy

**Handwriting Examination: Moving
from Art to Science**

K.H. Jarman
R.C. Hanlen

P.A. Manzollilo

RECEIVED
APR 26 1999
OSTI

April, 1999



Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RLO 18

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC06-76RLO 1830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831;
prices available from (615) 576-8401.

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161



This document was printed on recycled paper.

(9/97)

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Handwriting Examination: Moving from Art to Science

K.H. Jarman

R.C. Hanlen

P.A. Manzollilo

April 1999

Prepared for
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest National Laboratory
Richland, Washington 99352

Contents

List of Tables	iv
List of Figures	v
Acknowledgements	vi
Executive Summary	vii
1.0 Introduction	1.1
2.0 Data Description	2.1
2.1 Pilot Study	2.1
2.1.1 Handwriting Samples	2.2
2.1.2 Measurements and Characteristics	2.2
2.2 Primary Study	2.3
2.2.1 Handwriting Samples	2.3
2.2.2 Measurements and Characteristics	2.4
3.0 Handwriting Examination	3.1
3.1 Background	3.1
3.2 FDE Handwriting Examination Procedure	3.2
3.3 Quantitative Handwriting Examination	3.5
3.3.1 Constructing and Visualizing Handwriting Profiles	3.6
3.3.2 Cluster Analysis of Handwriting Profiles	3.9
4.0 Results and Discussion	4.1
5.0 Conclusions	5.1
6.0 References	6.1
Appendix A	A.1
Appendix B	B.1
Appendix C	C.1

List of Tables

1.	Height Ratios for Letters in Pilot Study	2.3
2.	Number of Handwriting Samples in Primary Study Grouped by Age and Sex	2.3
3.	Height Ratios for Letters in Primary Study.	2.4
4.	Results of Cluster Analysis of Primary Study Data	4.1
5.	Results of Cluster Analysis of Pilot Study	4.3
6.	Results of Simulated Grouping of Handwriting Profiles for $\alpha=0.2$	C.4

List of Figures

1.	Approach for Quantitative Handwriting Examination	3.6
2.	Box Plots for Five Exemplars from Primary Study	3.7
3.	Profile Centroid for Five Exemplars from Primary Study	3.8
4.	Profile Variation for Five Exemplars from Primary Study	3.9
5.	Profile Centroid for Three Exemplars from Primary Study	3.11
6.	Illustration of Grouping of Handwriting Profiles	4.2
7.	Truth Group Centers used in Handwriting Profile Cluster Validation	C.5

Acknowledgements

The authors would like to thank the US Postal Inspection Service Forensic Laboratories in Dulles, VA, New York, NY, Chicago, IL and San Francisco, CA and the Georgia Bureau of Investigation for providing the exemplars used in the pilot and primary studies. The authors would like to thank Heather Carlson and Charlotte Ware for their time and assistance with data collection and entry. In addition, the authors wish to thank Grant Sperry and Bob Muehlberger of the USPIS Forensic Laboratory for their input, time, and patience.

Funding for this effort was provided by the Technical Support Working Group under Government Order N3999899GO10799, under PNNL No. 28218/RD3802. The authors wish to express their gratitude for the support of Mr. Todd Anderson (TSWG) and Ms. Carrie Whitcomb (USPIS).

The work reported on here is part of a collaborative effort between staff of the United States Postal Inspection Service (USPIS) Forensic Laboratory, Memphis, Tennessee; Drexel University Data Fusion Laboratory (DFL), Philadelphia, Pennsylvania; and Pacific Northwest National Laboratory (PNNL), Richland, Washington. The authors also acknowledge the support of the forensic document examination community for their participation in the survey.

Executive Summary

The scientific basis for handwriting individuality and the expertise of handwriting examiners has been questioned in several court cases and law review articles. The criticisms were originally directed at the proficiency and expertise of forensic document examiners (FDE's). However, these criticisms also illustrate the lack of empirical data to support and validate the premises and methodology of handwriting examination. As a result, the admissibility and weight of FDE testimony has been called into question. These assaults on the scientific integrity of handwriting analysis have created an urgent need for the forensic document examination community to develop objective standards, measurable criteria, and a uniform methodology supported by properly controlled studies that evaluate and validate the significance of measurable handwriting characteristics.

FDEs, through a combination of training and experience, identify the important characteristics of a writing sample, conduct comparative examinations of habit agreement/disagreement, and assess the significance of the comparison. Some commonly compared characteristics include size, slant, line quality, character formation, height relationships of character forms, variation in height relationships and characteristic forms, and writing pressure. Examinations are conducted and conclusions are derived from experience and training with little, if any, empirical validation to support the premises or methodology. The scientific reliability of handwriting identification must be established through empirical validation of the premises and standardization of methodology.

In this document, we present a method for validating the premises and methodology of forensic handwriting examination. This method is intuitively appealing because it relies on quantitative measurements currently used qualitatively by FDE's in making comparisons, and it is scientifically rigorous because it exploits the power of multivariate statistical analysis. This approach uses measures of both central tendency and variation to construct a profile for a given individual. (Central tendency and variation are important for characterizing an individual's writing and both are currently used by FDE's in comparative analyses). Once constructed, different profiles are then compared for individuality using cluster analysis; they are grouped so that profiles within a group cannot be differentiated from one another based on the measured characteristics, whereas profiles between groups can. The cluster analysis procedure used here exploits the power of multivariate hypothesis testing. The result is not only a profile grouping but also an indication of statistical significance of the groups generated.

1.0 Introduction

The scientific basis for handwriting individuality and the expertise of handwriting examiners has been questioned in several court cases and law review articles. The criticisms were originally directed at the proficiency and expertise of forensic document examiners (FDE's). However, these criticisms also illustrate the lack of empirical data to support and validate the premises and methodology of handwriting examination. As a result, the admissibility and weight of FDE testimony has been called into question. These assaults on the scientific integrity of handwriting analysis have created an urgent need for the forensic document examination community to develop objective standards, measurable criteria, and a uniform methodology supported by properly controlled studies that evaluate and validate the significance of measurable handwriting characteristics.

A person's handwriting is the result of a continuous process that starts in grade school with the imitation of penmanship models, and develops into complex habitual motor skills. Handwriting habits are acquired and developed over a period of time and are influenced by a number of factors: perception, coordination, dexterity, muscle development, writing systems studied, family associations, occupation, education, cultural influences. These factors assert themselves through class and individual handwriting characteristics, which cumulatively compose a person's writing habits. It is the manifestation of these habits that enables handwriting identification.

Two fundamental assumptions provide the foundation for any handwriting examination: (1) an individual has some handwriting characteristics (i.e. habits) appearing consistently throughout his or her natural writing, and (2) the combination of all handwriting characteristics for an individual are unique. From these two assumptions, it follows that any natural writing may be associated with its author provided the writing contains sufficient individual habit features, and provided a fully comparable known writing sample is available. FDEs have accepted these concepts because they are reasonable rather than through scientific validation. Reliable handwriting identification capabilities have traditionally been achieved through FDE training and experience and not objective statistical analysis.

FDEs, through a combination of training and experience, identify the important characteristics of a writing sample, conduct comparative examinations of habit agreement/disagreement, and assess the significance of the comparison. Some commonly compared characteristics include size, slant, line quality, character formation, height relationships of character forms, variation in height relationships and characteristic forms, and writing pressure. Examinations are conducted and conclusions are derived from experience and training with little, if any, empirical validation to support the premises or methodology. The scientific reliability of handwriting identification must be established through empirical validation of the premises and standardization of methodology. The few studies that have approached the problem are deficient in many different areas: 1) studies are based on subjective or non-quantitative evaluations and classifications, 2) the experimental methodology lacks validation, 3) results are limited to observations and descriptions, and do not include statistical analysis, 4) analysis is limited to a few characteristics and habit areas, 5) a limited number of samples are used, 6) a lack of uniformity in samples limits cross comparability and analysis, and 7) publications and peer reviewed methods and results are lacking.

In this document, we present a method for validating the premises and methodology of forensic handwriting examination. This method is intuitively appealing because it relies on quantitative measurements currently used subjectively by FDE's in making comparisons, and it is scientifically rigorous because it exploits the power of multivariate statistical analysis. This

approach uses measures of both central tendency and variation to construct a profile for a given individual. (Central tendency and variation are important for characterizing an individual's writing and both are currently used by FDE's in comparative analyses). Once constructed, different profiles are then compared for individuality using cluster analysis; they are grouped so that profiles within a group cannot be differentiated from one another based on the measured characteristics, whereas profiles between groups can. The cluster analysis procedure used here exploits the power of multivariate hypothesis testing. The result is not only a profile grouping but also an indication of statistical significance of the groups generated.

The next section provides a description of the data available for this study. Section 3 contains a general description of the handwriting examination problem, and describes the approach developed here. Section 4 presents results and discussion of the statistical analysis of handwriting data. Finally, Section 5 presents conclusions and recommendations for future work.

2.0 Data Description

Measurements were made from selected words appearing on different handwriting exemplars. The measurements were made using a mylar sheet containing a grid made up of 0.2 cm boxes placed over the exemplar. Therefore, all measurements are reported in centimeters.

Measurements are available from two separate studies, a pilot study and a primary study. The pilot study data set consists of 130 exemplars collected by the US Postal Inspection Service in Memphis, Tennessee. The primary study data set consists of 500 exemplars collected by the Georgia Bureau of Investigation. All handwriting forms were collected for the purpose of comparison to questioned writing during investigations. Subjects providing exemplars included both suspects and victims. These exemplars were collected over approximately a 20 year time period under varied conditions that are not explicitly known.

The types of measurements made include the following:

- Total height – height of baseline to apex of characters above baseline.
- Upper portion – height of apex of bulb, hump, or point of intersection to apex of a character above baseline.
- Lower portion – height of baseline to apex of bulb, hump, or point of intersection of a character above baseline and/or height of baseline to apex of bulb, hump, or point of intersection of a character below the baseline
- Length of text – the total length of a word or string of characters.
- Length of line – the total length of a printed line.
- Number of characters – total number of characters actually written including length one character or less open spaces.
- Open space – total length of two or more open spaces between words of text.

The measurements were chosen based upon their frequency of occurrence in the form, limited amount of subjective assessment necessary and use in traditional handwriting examination.

Only height relations are used for quantifying handwriting examination here. To simplify notation, the measured height of a character is denoted by the character in italics, and the height of a portion of a character is denoted by the character followed by a further descriptor. For example, the measurement denoted *F* refers to the total height of the character "F", and *F-lower* refers to the lower portion height of the character "F". Similarly, the measurement denoted *t* refers to the height of the character "t" and *t-x* refers to the height of the "t" crossing from the baseline. All other types of measurements are denoted more explicitly.

2.1 Pilot Study

The purpose of this study was to determine if objective measurements could be used to study the significance of specific handwriting characteristics and further to examine the axiom of handwriting individuality using the cumulative significance of these limited measurable handwriting characteristics. Measurements were made during over a two month period at the end of 1996 for presentation of preliminary results at the annual American Academy of Forensic Sciences meeting in February 1997. The handwriting samples were collected over a number of years prior to 1996. The handwriting samples used in this study were selected for a variety of reasons: 1) the samples were extensive text encompassing numerous habit areas, 2) handwriting habit areas were repeated throughout the text allowing variation to be assessed, 3) the samples were in a standardized format allowing direct comparisons between and within writers, 4) the baseline provided a fixed endpoint for measurements limiting subjectivity, and 5) these samples

were a subset of a large group of similar forms completed by over 700 individuals. The limitations included the fact that these samples were not original handwriting (all samples were photocopies), all samples were collected at a single sitting, the conditions under which the form was completed was unknown and little if any demographic data was available on the writers. To study the ability of the measurements to characterize the handwriting of an individual, 5 samples were measured twice and recorded as two different samples. To study the effect of unnatural handwriting on the measurements, two individuals prepared multiple forms emphasizing unnatural handwriting relationships in the areas to be measured these comprised another 5 samples of the pilot study. The precision of the method and the analyst was checked by measuring one form repeatedly throughout the two month period.

2.1.1 Handwriting Samples

Very limited data (name and city of residence) is available on the subjects who provided the handwriting samples that were used in the pilot study. 130 handwriting exemplars from 122 individuals (61 males and 61 females) were collected using U.S. Postal Service Inspection Service Form 582. Form 582 is included in Appendix A. All samples were evaluated and selected only if the handwriting appeared natural (except where unnatural handwriting exemplars were created specifically for the purposes of the study).

2.1.2 Measurements and Characteristics

For each of 130 exemplars collected, 93 measurements were made. These measurements were then combined to give a total of 80 ratios falling into four categories: (1) characteristics of the pair "th", (2) characteristics of numbers, (3) height relationships of letters within words, and (4) lateral spacing of characters and lines. A description of the measurements and characteristics for each category follows.

All six occurrences of the pair "th" appearing in the exemplar were measured. These occurrences are outlined with boxes in Appendix A. The heights t and h were measured, along with the heights $t-x$ and the h -hump. For each occurrence of the pair, these four measurements were combined in the form of the following ratios: t/h , $t-x/h$, h -hump/ t , h -hump/ h , and $t-x/t$.

All occurrences of the numbers "3", "4", and "9" were measured, and are marked by boxes on the exemplar in Appendix A. For each occurrence of these numbers in the exemplar, the total height and upper portion height were measured. These measurements were then combined to give the following ratios: 3 -upper/ 3 , 4 -upper/ 4 , 9 -upper/ 9 , $3/4$, $3/9$, and $4/9$.

Height relationships were measured from the words highlighted on the exemplar in Appendix A. The letters measured include any occurrences of the following letters in six selected words: "E", "F", "N", "K", "L", "U", "d", "I", "t", "b", "g", "y". For each word, the capital letter was compared to all of the selected lower case letters appearing in the same word. In addition, letters with upper and lower extenders (including uppercase letters) were compared to their respective portions, other letters in the same word, and extender letters between words. These comparisons were made using ratios of measured heights. Some of the measurements and the various between-letter, within-word ratios are given in Table 1. Note that when more than one occurrence of a letter appears in a word, the first occurrence is denoted by a subscript 1, the second by a subscript 2, etc

Table 1. Height Ratios for Letters in Pilot Study

<i>Word/Sequence</i>	<i>Measurements Made</i>	<i>Ratio/Characteristic</i>
Edward	<i>E, d₁, d₁-bulb, d₂, d₂-bulb</i>	<i>d₁/E, d₁-bulb/E, d₂/E, d₂-bulb/E</i>
Fallett	<i>F, l₁, l₂, t₁, t₁-x, t₂, t₂-x</i>	<i>l₁/F, l₂/F, t₁/F, t₁-x/F, t₂/F, t₂-x/F</i>
Newburg	<i>N, b, b-bulb, g, g-bulb</i>	<i>b/N, b-bulb/N, g/N, g-bulb/N, b/g-ext, b-bulb/g-ext</i>
Kitty	<i>K, t₁, t₁-cross, t₂, t₂-cross, y, y-upper</i>	<i>t₁/K, t₁-x/K, t₂/K, t₂-x/K, y/K, y-upper/K, t₁/y-ext, t₁-x/y-ext, t₂/y-ext, t₂-x/y-ext</i>
Langdon	<i>L, g, g-bulb, d, d-bulb</i>	<i>g/L, g-bulb/L, d/L, d-bulb/L</i>
Ulysses	<i>U, l, y, y-upper</i>	<i>l/U, y/U, y-upper/U</i>

The lateral spacing measurements were taken from the three lines of the exemplar in Appendix A labeled "lateral". In each case, the following measurements were made: length of text, length of line, number of characters, and open space. These measurements were then combined to give the following ratios: length text/length line, length text/# characters, and open space/length text.

2.2 Primary Study

These measurements were originally made for the purpose of studying the use of objective measurements to evaluate limited handwriting characteristics and variation as it relates to age and gender. The measurements were made during a two month time period in 1997. The handwriting samples were collected during 1981 through 1985. The handwriting samples used in this study were selected for a variety of reasons: 1) the samples were extensive text encompassing numerous habit areas, 2) handwriting habit areas were repeated throughout the text allowing variation to be assessed, 3) the samples were in a standardized format allowing direct comparisons between and within writers, 4) the baseline provided a fixed endpoint for measurements limiting subjectivity, and 5) these samples were a subset of a large group of similar forms completed by over 2900 individuals. The precision of the method and the analyst was checked by measuring one form repeatedly throughout the two month period. The limitations of these exemplars include: 1) all samples were collected at a single sitting, 2) the conditions under which the form was completed was unknown and 3) the geographic distribution of the subject writers may be concentrated in the southeastern United States. Subjects in the primary study include 240 women between the ages of 15-55 and 260 men between the ages of 14-59.

2.2.1 Handwriting Samples

The exemplar given in Appendix B was used for collecting the handwriting samples in this study. 500 handwriting samples (a subset of a larger group of approximately 2900 samples) were collected and grouped by age and sex. For each group, the number of samples collected is provided in Table 2.

Table 2. Number of Handwriting Samples in Primary Study Grouped by Age and Sex

	<i>Age 10-19</i>	<i>Age 20-29</i>	<i>Age 30-39</i>	<i>Age 40+</i>
Female	70	70	70	30
Male	70	70	70	50

2.2.2 Measurements and Characteristics

All measurements described below were taken from each exemplar except when a character was indiscernible or missing. Out of 500 exemplars collected, 88 have one or more missing measurement. For each exemplar, a total of 56 measurements were made and combined into 78 ratios. All of the measurements used fall into the category of character height. The words from which measurements were collected are highlighted in Appendix B.

Fourteen words each containing a capital letter, a lowercase "a" and one or more lowercase characters with an upper extender were selected on each form. For each of 14 words, the height of the capital letter was measured along with total height and lower portion of any occurrences of the following: "a", "d", "h", "l", and "b". Within each selected word, the measurements were combined in the form of ratios between the capital letter and the various lower case letters, between the *a* and the lower case consonants, and between the *l* and other lower case consonants. The ratios computed for each word are presented in Table 3.

Table 3. Height Ratios for Letters in Primary Study

Word	Measurements Made	Ratio/Characteristics
Edward	<i>E, d₁, d₁-hump, a</i>	<i>E/d₁, E/d₁-hump, E/a, a/d₁, a/d₁-hump</i>
Micheal	<i>M, h, h-hump, a, l</i>	<i>M/h, M/h-hump, M/a, M/l, a/h, a/h-hump, a/l, l/h, l/h-hump</i>
Charles	<i>C, h, h-hump, a, l</i>	<i>C/h, C/h-hump, C/a, C/l, a/h, a/h-hump, a/l, l/h, l/h-hump</i>
William	<i>W, l, a</i>	<i>W/l, W/a, a/l</i>
Raymond	<i>R, a, d, d-hump</i>	<i>R/a, R/d, R/d-hump, a/d, a/d-hump</i>
Taylor	<i>T, a, l</i>	<i>T/a, T/l, a/l</i>
Donald	<i>D, a, l, d, d-hump</i>	<i>D/a, D/l, D/d, D/d-hump, a/l, a/d, a/d-hump, l/d, l/d-hump</i>
Thomas	<i>T, h, h-hump, a</i>	<i>T/h, T/h-hump, T/a, a/h, a/h-hump</i>
Lamb	<i>L, a, b, b-bulb</i>	<i>L/a, L/b, L/b-bulb, a/b, a/b-bulb</i>
Samuel	<i>S, a, l</i>	<i>S/a, S/l, a/l</i>
Elizabeth	<i>E, l, a, h, h-hump</i>	<i>E/l, E/a, E/h, E/h-hump, a/l, a/h, a/h-hump, l/h, l/h-hump</i>
Vaughn	<i>V, a, h, h-hump</i>	<i>V/a, V/h, V/h-hump, a/h, a/h-hump</i>
Franklyn	<i>F, a, l</i>	<i>F/a, F/l, a/l</i>
Upshaw	<i>U, h, h-hump, a</i>	<i>U/h, U/h-hump, U/a, a/h, a/h-hump</i>

3.0 Handwriting Examination

In this section, we present an approach for quantifying the handwriting examination procedure. In the following two sections, we provide some background into handwriting examination by summarizing the assumptions and procedures currently used. We discuss issues in associating an unknown source exemplar with a known source exemplar. In Section 3.3., we present an approach for quantifying handwriting characteristics, constructing handwriting profiles, and comparing profiles for different individuals in a way that takes statistical uncertainty into account.

3.1. Background

There are two fundamental assumptions that provide the foundation for any handwriting examination:

- (1) an individual has some identifiable set of handwriting characteristics (i.e. habits) appearing consistently throughout his or her natural writing, and
- (2) the combination of all handwriting characteristics for an individual are unique.

A person's handwriting is the result of a continuous process that starts in grade school with the imitation of penmanship models, and develops into complex habitual motor skills. For a given individual, these motor skills manifest themselves as handwriting characteristics (see, e.g. Moenssens, et.al (1995)), so assumption (1) is universally accepted. However, the questioned document community has recently been criticized on the basis of assumption (2) as well as on the ability of handwriting examiners to accurately associate or differentiate questioned and known source samples (see, e.g. Starrs (1995)).

Practically speaking, assumption (2) is impossible to prove. Osborn (1944) has provided an argument for the individuality of handwriting, and a few authors (Huber (1990), Huber and Headrick (1989), Blake (1988)) have proposed approaches for addressing the criticisms arising from assumption (2). In particular, all that is required to make strong statements about the identity of the author of a questioned document is a demonstration that the combination of characteristics exhibited in the questioned document are consistent with known natural writings and are sufficiently rare to rule out a large percentage of the population. As a result, a number of studies have been conducted which examine the frequency of various characteristics within a sample population (Horan (1993), Horton (1996), McClary (1997), Horan and Horan (1984), Zimmerman (1990), Muehlberger, et al (1977)). A few authors have proposed methods for statistically estimating the rarity of an individual's handwriting based on a set of characteristics (Hilton (1995), Rhodes (1997)). These approaches involve the following steps:

- (a) identify a characteristic (such as letter slant)
- (b) categorize the characteristic into a few levels (such as forward, vertical, or backward slant)
- (c) quantify the relative frequency of occurrence of each level for each characteristic, and
- (d) multiply the relative frequencies of each observed characteristic to estimate the rarity of a set of characteristics.

This approach suffers some limitations. First, violation of an assumption implicit in step (d) can result in grossly inaccurate estimates of the relative frequency of a set of handwriting characteristics. In particular, the relative frequency of a set of characteristics can be accurately estimated by multiplying individual relative frequencies only when the characteristics of interest are statistically independent. This means that the specific level of a given characteristic has no impact on the levels of any other characteristic. As an illustration, characteristics that violate this

assumption of independence might include extender on the letter "d" and beginning strokes of the letter "r". An open loop on the "d" extender and an "r" that begins like an "n" were both taught uniquely by the Palmer system of writing at one point. Therefore, it is possible that the observance of an open loop on the "d" could result in a higher than average probability of also observing an "r" that begins like an "n". Using the above approach when characteristics are not statistically independent of one another will have serious implications because it can result in an estimate of the rarity of an individual's habits that is much smaller than the actual value. For a further discussion of this issue, the reader is referred to Hilton (1995) and Totty (1985).

Another issue regarding this approach involves the manner in which characteristics are quantified and used. Information is lost in categorizing characteristics as in (b) since a given characteristic can generally have a whole continuum of values rather than a few discrete ones. For example, suppose one of the characteristics of interest is whether or not the letter "a" is open or closed. In reality, the "a" can range from being fully closed (gap of zero), to wide open. An individual may exhibit some characteristic gap size, but this gap size can vary greatly. Therefore, not only the absolute gap size but also the variation in gap size can be considered individual characteristics. A study that forces the examiner to make a potentially arbitrary and limited determination about whether the "a" is open or closed doesn't take a writer's variation into account. Therefore, information that could be crucial to making a determination is lost in comparing characteristics in this manner. In comparative analyses, FDE's consider variation, so it is important that any quantitative method for comparing handwriting also takes variation into account.

Any method for quantifying handwriting examination should incorporate appropriate statistical sampling techniques. For the above approach to be accurate, a random sample of data representative of the population from which questioned documents are drawn must be used to quantify the relative frequencies of each characteristic. Obtaining such a sample can be a very cumbersome process, especially when the characteristics of interest are rare in the population. In past studies, people have either neglected this issue or made simplifying assumptions that the relative frequency of each level for a given characteristic is the same (e.g. 1/3 each for forward, vertical, and backward slant). By neglecting statistical sampling issues, the estimated relative frequency of a set of characteristics can be grossly inaccurate, particularly when specific habits are prevalent in one group of the population but rare in another. Assuming the relative frequency of different levels of a characteristic are equal is convenient, but not necessarily accurate and should therefore be avoided. The data available for the research presented here does not represent a random sample, but was convenient to obtain. It is therefore used solely for the purpose of developing a methodology that can then be used on data collected from an appropriate random sample at a later date. The reader is warned that the results presented here should not be generalized, but rather should be considered a demonstration of the capabilities of the methods developed in this research.

3.2. FDE Handwriting Examination Procedures

The examination and comparison of handwriting is a multistage process. The process involves independent examinations of 1) the questioned writing and 2) the known writing, followed by 3) comparison of the questioned with the known. This process evaluates both the specific handwritten entries and the entire document for all relevant information.

The examination begins with an overall survey of the documents and writing to be compared. The examiner first determines if the questioned and known writings are comparable. (i.e. handwriting to handwriting or hand printing to hand printing) If the questioned and known writings have limited or no comparability, the examination may end at this point.

The examination of the questioned writing proceeds from the general to the specific. The questioned is examined to answer the following questions, if possible:

- 1) How was the questioned writing produced?
- 2) Is the questioned writing original? (i.e. as opposed to an office machine or carbon copy)
- 3) Is the questioned writing freely and naturally written? (Characteristics used to assess the naturalness of the writing can include line quality, speed, rhythm, and pressure variation.)

If the examination reveals potentially limiting factors such as an absence of original writing, unnaturalness or distortion, these factors are considered throughout the examination as they may impact the examination and findings.

The questioned writing is then examined to discover the specific features and characteristics that comprise the individual writing habits of the writer(s). This process includes an examination of the questioned writing for evidence of multiple writers. The following steps outline some (but not all) of the handwriting features and characteristics that are examined.

- 1) The examination can begin with the more general features of the questioned writing such as overall size, spacing, slant, baseline habit and formatting (margins, punctuation). The overall line quality, rhythm, relative skill level and speed are also examined.
- 2) Next the less conspicuous or more specific features and characteristics of the questioned writing are examined. These features can include: letter design, stroke direction, pressure variation, angularity, connecting strokes, relative size, spacing, slant and baseline habits, hesitations, pen lifts, disconnects diacritics, flourishes, embellishments and idiosyncrasies.
- 3) The variation of each feature or characteristic observed in the questioned writing is assessed where possible.
- 4) The examiner evaluates the significance of the writing features and characteristics and their demonstrated range of variation observed in the questioned writing. The training and experience of the examiner plays an important role in this evaluation. The cumulative significance of all the writing habits is used to determine if the questioned writing is identifiable and of sufficient quantity.

The known writing is now examined utilizing the same procedure as the questioned writing. An extensive examination of the writing features and characteristics demonstrated in the known writing allow the examiner to evaluate the known writer's range of writing variation.

The final stage of the examination is the side by side comparison of the questioned and known writing. The following steps included in this stage:

- 1) A comparison of the writing habits observed in the questioned and the known.
- 2) An evaluation of the significance of environmental factors such as: writing conditions, relative dates of preparation, writing instrument, paper and writing surface on the questioned and known writing.
- 3) An evaluation of the significance of fundamental differences, dissimilarities, absent characteristics, accidentals, similarities and variation observed in the questioned and the known writing. These are evaluated both individually and cumulatively.

At the conclusion of the examination, the examiner expresses his/her conclusion as to whether or not the questioned and the known writing were written by the same person. The following conclusions can be expressed:

- 1) Identification – a definitive conclusion that the writer of the known writing wrote the questioned writing. An absence of fundamental differences or unresolved discrepancies and the presence of substantial significant individual similarities in the questioned and the known writing are required for this conclusion.
- 2) Probably Did Write – a less than definitive conclusion that the writer of the known writing probably wrote the questioned writing. An absence of fundamental differences and the presence of many significant individual similarities in the questioned and the known writing are required for this conclusion. The less than definitive nature of this finding may be the result of limiting factors such as absent characteristics, and/or the quality or quantity of either the questioned or known writing.
- 3) Indications Did Write – a less than definitive conclusion that there are indications the writer of the known writing wrote the questioned writing. An absence of fundamental differences and the presence of some significant individual similarities in the questioned and the known writing are required for this conclusion. The less than definitive nature of this finding may be the result of significant limitations such as the cumulative significance of the similarities, an absence of fully comparable writing, disguise or distortion, and/or the quality or quantity of either the questioned or known writing.
- 4) No Conclusion – is a less than definitive finding that the writer of the known could not be identified or eliminated as the writer of the questioned writing. This conclusion may be the result of an absence of comparable writing or of the questioned having little or no value for identification.
- 5) Indications Did Not Write – is a less than definitive finding that there are indications the writer of the known writing did not write the questioned writing. The presence of some significant dissimilarities in the questioned and the known writing is required for this conclusion. Similarities may also be present in the questioned and the known writing. The less than definitive nature of this finding may be the result of significant limitations such as the cumulative significance of the similarities, an absence of fully comparable writing, disguise and/or the quality or quantity of either the questioned or known writing.
- 6) Probably Did Not Write - is a less than definitive finding that the writer of the known writing probably did not write the questioned writing. The presence of many significant dissimilarities in the questioned and the known writing is required for this conclusion. Similarities may also be present in the questioned and the known writing. The less than definitive nature of this finding may be the result of limiting factors such as the cumulative significance of the dissimilarities and/or the quality or quantity of either the questioned or known writing.
- 7) Elimination – this is a definitive conclusion that the writer of the known writing did not write the questioned writing. The presence of fundamental differences between the handwriting features and characteristics of the questioned and the known writing are required for this conclusion. Similarities may also be present in the questioned and the known writing.

3.3. Quantitative Handwriting Examination

In the research presented here, we have developed a methodology for characterizing and assessing the individuality of a person's handwriting that circumvents the difficulties inherent in the methods summarized in Section 3.1. In particular, we propose a method for characterizing handwriting that takes statistical dependence between characteristics into account, quantifies the central tendencies and variation in characteristics for a given individual, and exploits the fact that different individuals exhibit different central tendencies and variation in their handwriting. We use the estimated mean and variance of different measured characteristics to construct a *handwriting profile*. We then compare handwriting profiles of different individuals and group them so that profiles within a group cannot be statistically differentiated from one another, whereas profiles between groups can.

This approach is conceptually illustrated in Figure 1. In particular, an individual's handwriting profile can be constructed from the estimated mean and variation of a combination of measured characteristics comprising his or her writing. It is assumed that an individual's handwriting profile falls into one of a number of groups. Each group is made of all profiles that cannot be differentiated from one another based on the measured characteristics. Any unknown source exemplar can then be characterized by its profile and placed into one of the groups using a statistical comparison. Association of an unknown source exemplar to an individual can then be made by examining the size of its group relative to the population (or the rarity of a given profile type within the population). In Section 3.3.2., we will describe formation of these profile groups and discuss how membership of a profile in a group actually corresponds to a degree of association which takes statistical uncertainty into account.

As an example, suppose questioned and known source handwriting documents both fall into group B with some high degree of association. Then it can be said with a high level of confidence that the questioned document and the known source exemplar were both written by an individual in profile group B. Since group B contains some percent of the population (3% in this example), the probability that a randomly chosen individual fits into group B is 0.03, or 3 in 100.

The approach for quantifying handwriting requires three stages. The first stage includes constructing and visualizing a profile for an individual's handwriting based on a set of measurements. The second stage involves constructing groups of handwriting profiles such that: (a) different groups of profiles have statistically different handwriting characteristics, and (b) profiles within a group cannot statistically be differentiated from one another based on the measured characteristics. This type of analysis is called *cluster analysis*. The third stage involves estimating the percentages of the population falling into each group. Note that the third stage is not addressed here because of the lack of an appropriate random sample of data. However, this stage could easily be addressed in future studies. In this section, we summarize the concepts behind the methods used here. Section 3.3.1 discusses the construction and visualization of handwriting profiles from measured data. Section 3.3.2. describes a method for obtaining groups of profiles. A complete mathematical description of the methods used here is given in Appendix C.

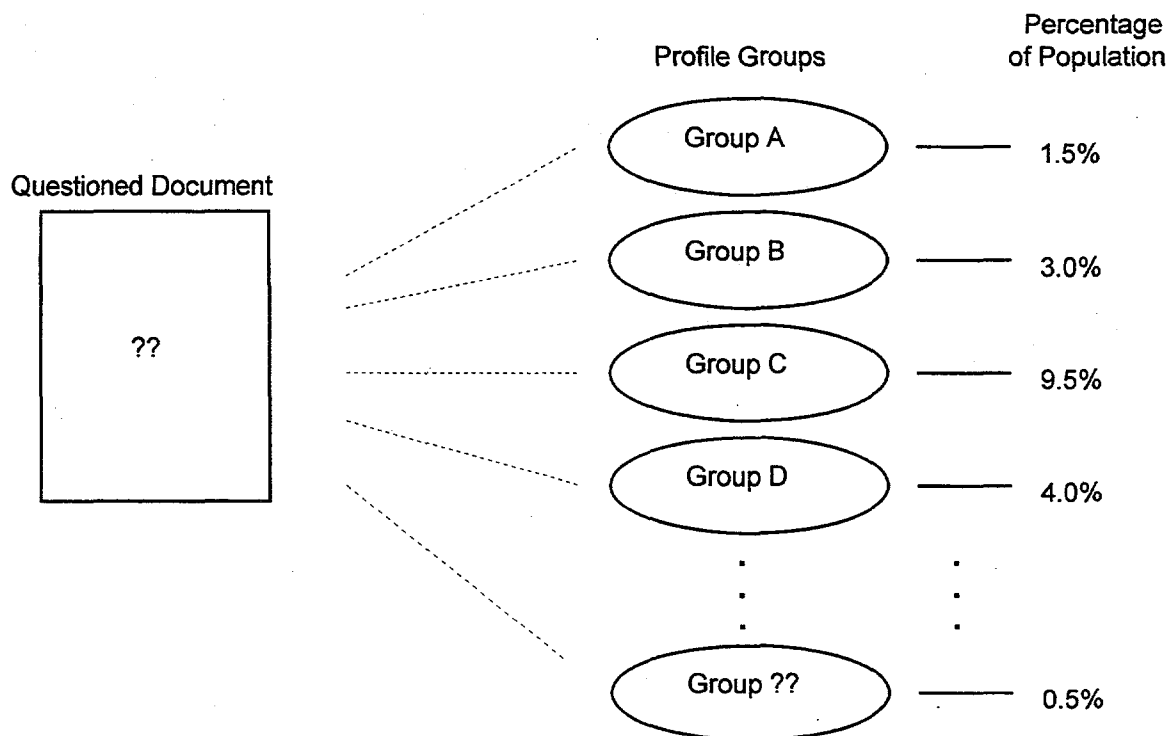


Figure 1. Approach for Quantitative Handwriting Examination. Profiles are grouped statistically so that profiles between groups can be differentiated from one another whereas profiles within a group cannot. Each group represents some percentage of the population, and the percentage represented by each group corresponds to the rarity of the habits comprising a handwriting profile. The strength with which a questioned document can be associated to a known source document is determined by (a) the statistical certainty that the questioned document belongs to the group containing the known source sample, and (b) the percentage of the population represented by that group.

3.3.1 Constructing and Visualizing Handwriting Profiles

Forensic document examiners use both central tendency and variation of exemplar(s) to characterize an individual's handwriting. Therefore, we construct a handwriting profile using the centroid, or average and variation, or variance of a set of characteristics. In the initial phase of this research, the only characteristics used are a limited number of character height ratios within a word. These height ratios include internal portion (ratio of the height of an extender letter to the height of the lower portion extending only to midline such as *b/b-bulb*), and external portion (ratio of the height of a capital letter to a lower case "a" such as *E/a*). By using ratios, the analysis presented here becomes insensitive to overall size of handwriting, and is therefore not limited to the specific exemplar forms used here.

One way to illustrate the centroid and variation of different measurements for a given exemplar is shown in Figure 2. In the figure, a boxplot of different height characteristics is plotted for five exemplars. Note that all measurements for a given exemplar have been divided by the average *l* height to remove sensitivity to overall writing size. For each height ratio, the center of the box represents the median, or middle value of the measurements while the box encloses the middle 50% of the observations (i.e. the upper and lower edges of the box represent the first and third

quartiles of the data, respectively). Finally, the lines, or whiskers indicate the location of the largest and smallest observations. Clearly, the center line in each box provides a rough representation of the central tendency of a given height relation. Furthermore, since the box for a given height relation encloses the middle 50% of the observations, the overall height of a box represents the variation for that measurement. A larger box indicates a height characteristic with larger variation.

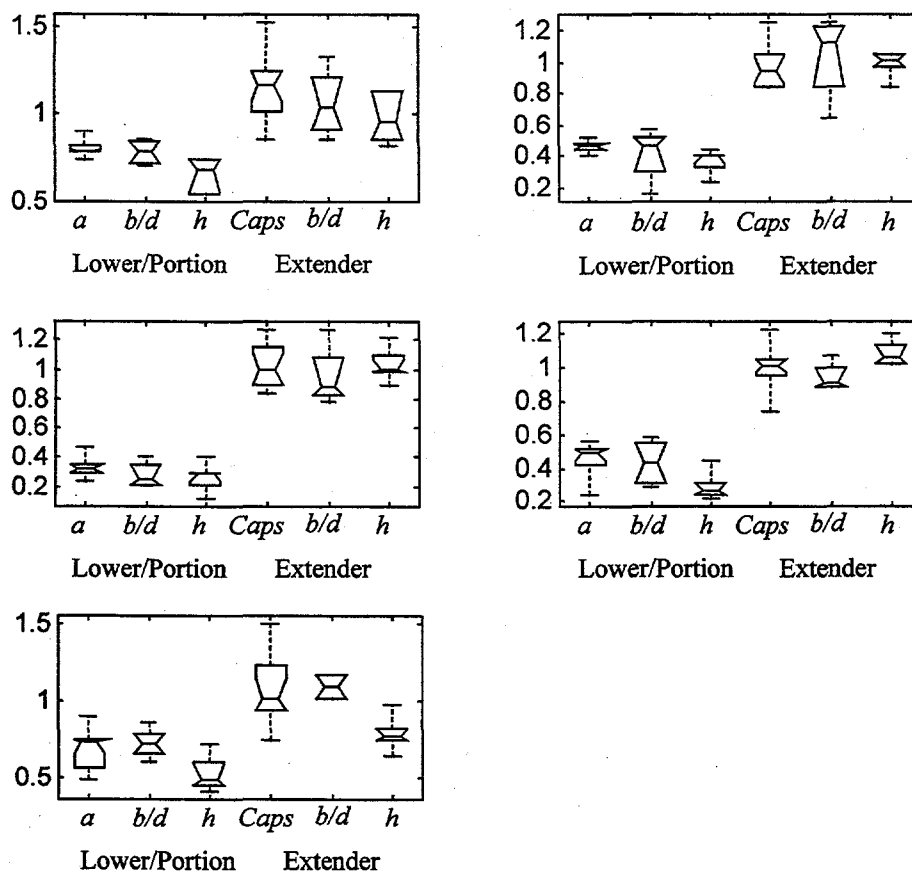


Figure 2. Boxplots for Five Exemplars from Primary Study. Each boxplot illustrates the central tendencies and variation of different characteristics for a given exemplar. The central tendency for a given characteristic is represented here by the center line in each box. The center line is the median, or middle value of the measurements. The variation is represented by the size of each box, which encloses the middle 50% of the measurements for a given characteristic. A larger box corresponds to larger variation.

By combining these individual measurements into height ratios, a profile for each exemplar can be constructed. Figures 3 and 4 illustrate the profiles for the five individuals whose measurements are displayed in Figure 2. In Figure 3, a profile is visualized by plotting either (a) capital letter height against the height of the lower case *a*, or (b) extender height against the corresponding lower portion height. Case (a) quantifies the across letter height relationship while case (b) quantifies the within letter height relationship. Note that by displaying the profile in this manner, the height ratios are not computed explicitly but are made implicitly by plotting one versus the other. This implicit method for displaying height relationships does not remove

sensitivity to overall writing size. Therefore, the measurements have been normalized through dividing by the average *l* height.

This method of displaying handwriting profiles is useful since profiles for several individuals can be displayed on the same plot and compared. In the figure, the average height for each exemplar is denoted by a large number, and the individual measurements are denoted by smaller numbers. The ellipse enclosing the measurements for a given exemplar represents an uncertainty region containing all the data and is useful for comparing the centroid of different exemplars. In particular, the extent to which two ellipses overlap indicates the extent of similarity in centroid between two exemplars.

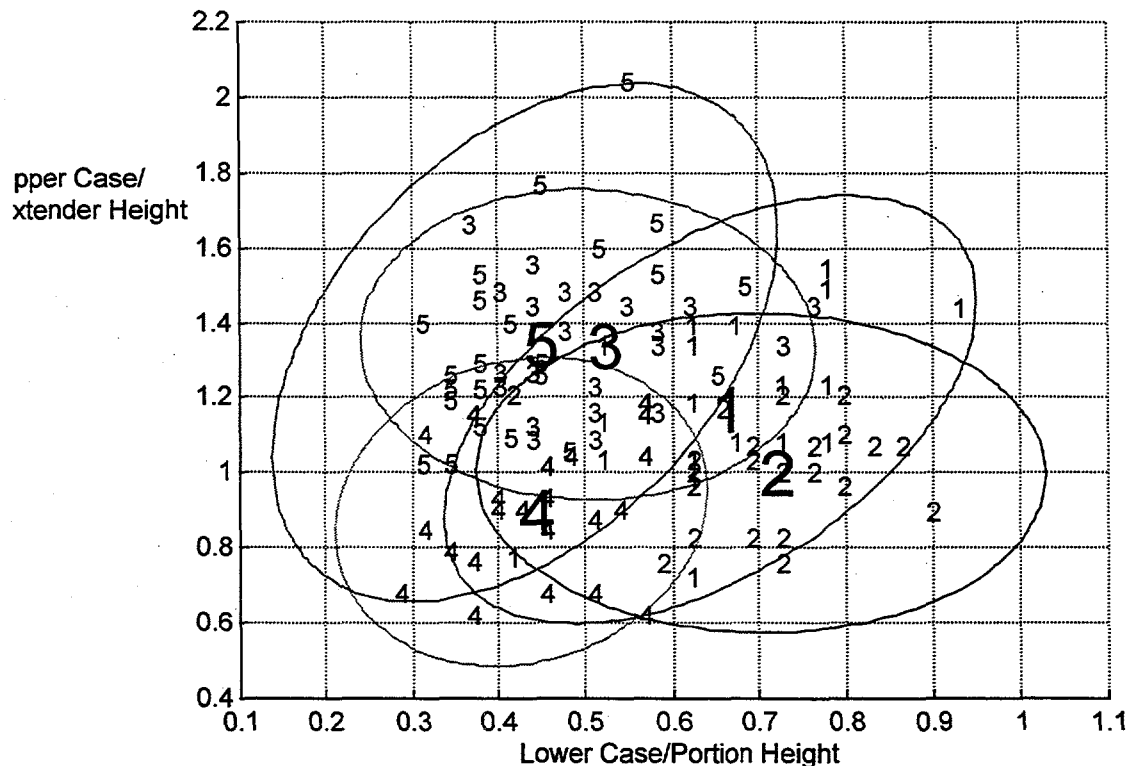


Figure 3. Profile Centroid for Five Exemplars from Primary Study. The y axis represents the upper case/extender height for a given characteristic while the x axis represents the corresponding *a* or lower portion height. The ellipses enclose the extent of the measurements for each exemplar. The extent to which the ellipses overlap corresponds to the degree of similarity in central tendency between exemplars.

Figure 4 illustrates variation of a handwriting profile by plotting the variance of the capital letter (extender) height against the variance of the lower case (lower portion) height. Analogous to Figure 3, average variance is denoted by a large number, and the variance of each characteristic is denoted by a smaller number. The ellipses enclose the extent of the variation. In this case,

similarity in variation can be determined by examining the extent of overlap between the variance profiles for different exemplars.

The large amount of overlap in the centroid exhibited in Figure 3 suggests that the central tendencies of the limited height ratios measured in the primary study do not give much information in distinguishing individuals' handwriting. However, the significantly smaller degree of overlap in Figure 4 suggests that variation in height ratios can be used to help distinguish different individual's handwriting. This is an important observation because it implies that the combination of central tendency and variation provide much more discriminatory information than central tendency alone, and should therefore be used in any quantitative analysis of handwriting data.

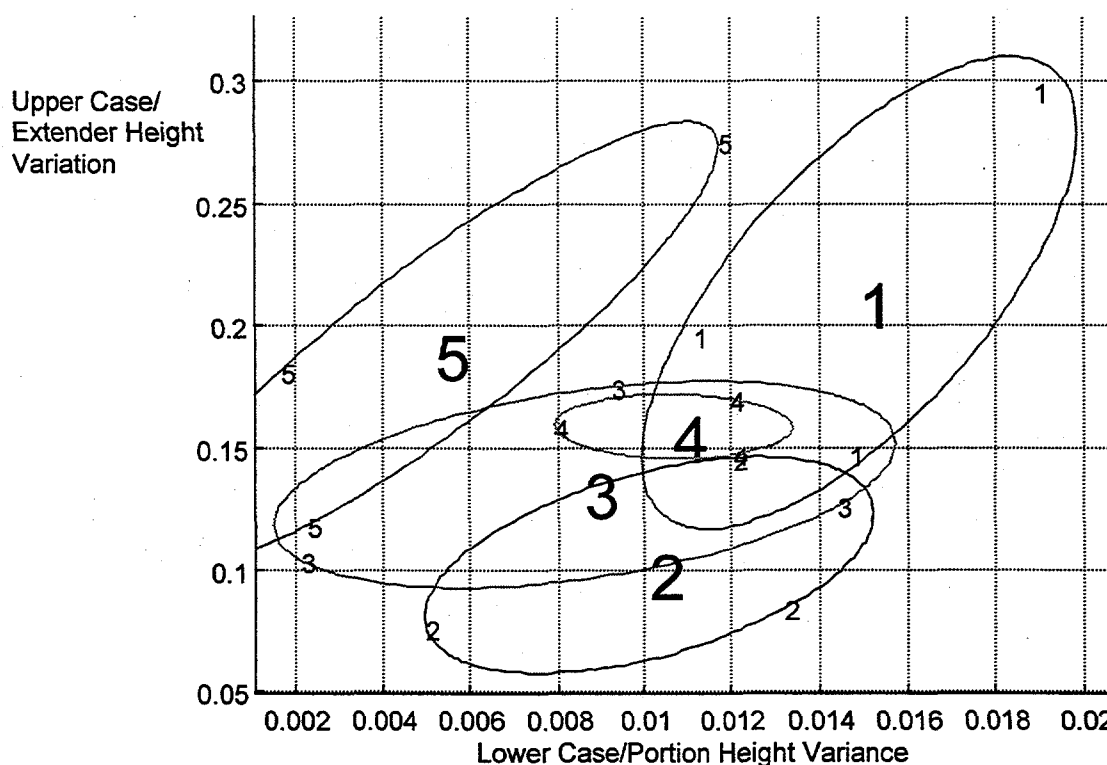


Figure 4. Profile Variation for Five Exemplars from Primary Study. The y-axis represents the variation in upper case/extender height and the x-axis represents the corresponding *a* or lower portion height. The ellipses enclose the extent of the variation. The degree to which two ellipses overlap corresponds to the degree of similarity in variation between two exemplars.

3.3.2. Cluster Analysis of Handwriting Profiles

In constructing groups of exemplars based on the characteristics studied here, it is useful to study Figures 3 and 4. For both centroid and variation, the ellipses enclosing the observations for a given exemplar represent uncertainty regions inside which the handwriting can vary. The extent to which two ellipses overlap corresponds to the extent of similarity or confusion between

the writing on the two exemplars. For example, if the ellipses for two profiles have no overlap in either centroid or variation, it is highly unlikely that observations from the two exemplars could be confused. In this case, the two profiles are easily differentiable. At the other extreme, if the centroid and variation ellipses for two profiles fall directly on top of one another, any measurement made on one exemplar is just as likely to have come from the other. Therefore, the two different exemplars are completely indistinguishable with respect to the measured characteristics. Note that the ellipses in Figures 3 and 4 represent a very limited amount of information – either centroid or variation in across or within letter height ratios. Since the data collected from an exemplar is limited, it comes as no surprise that several of the ellipses in Figures 3 and 4 overlap significantly.

The goal of this study is to combine information for both centroid and variation into a handwriting profile and perform a multivariate statistical analysis to determine the degree of individuality of each profile. This is performed by constructing groups of profiles where profiles in different groups can be statistically differentiated, but profiles within a group cannot. The groups are constructed by incorporating multivariate hypothesis tests into a cluster analysis procedure. The method used is an extension of a method called the Leader Algorithm (Hartigan (1975)) and proceeds as follows:

- (0) Decide what statistical significance level α is desired for each group. Take the first profile and construct a group containing only that profile. Set the group center equal to the first profile (mean and variance).
- (1) Select the next profile. Using statistical hypothesis tests compare this profile to each group center in sequence.
 - For a given group, if the profile matches the group center to a significance level α , add the profile to the group and recompute the group center as the average of all member profiles.
 - If the profile does not match any group center, construct a new group containing this profile, and set the group center equal to this profile (mean and variance).

Conceptually, this algorithm produces groups of profiles clustered around some group center with uncertainty region $1-\alpha$. An illustration of this is provided in the next section. For more detail, see Appendix C.

One way to approach this analysis is to assume that any two individual profiles are fundamentally different and determine what significance level α would be required to distinguish every exemplar. In terms of the algorithm presented above, this corresponds to increasing the value of α until the cluster analysis generates only groups with one member profile. This corresponds to determining how close an unknown source exemplar must be to its author's profile in order to make a positive match. This can be illustrated by examining Figure 5. In the figure, the centroids for three exemplars are plotted along with ellipses enclosing all of the measured data. The three profiles plotted are samples 165, 965, and 1186 from the primary study. The ellipses enclosing all the data can be thought of as uncertainty regions covering virtually all of the data (say, 99.9% or more) for a given individual. If we shrink the size of the ellipses, less overlap between individuals will occur, but the uncertainty regions will enclose a smaller percentage of the data. Indeed, if we make the ellipses small enough, then no two profiles' uncertainty regions will overlap and all of the profiles are unique.

Reducing the size of the ellipses increases our ability to differentiate between two exemplars from different individuals, however, it also reduces our ability to accurately match two

exemplars from the same individual. In the case of Figure 5, the smaller ellipses enclose 41% of the measurements. We could therefore say, loosely, that in looking at a single measurement from one of the authors in Figure 5, 41% of the time the measurement would fall inside the smaller ellipse and therefore be correctly associated with its author. The remaining 59% of the time, the measurement would fall outside the appropriate ellipse and no strong determination about authorship could be made. In this case, only degrees of association between the measurement and the three profiles could be made.

By taking multiple measurements from an unknown writing, the ability to correctly associate and differentiate with known source profiles can be strengthened. In particular, if enough measurements are collected in the case of Figure 5, we can be assured that the average of the measurements will be inside the appropriate ellipse so that a strong determination about authorship can be made. Unfortunately, however, the amount of unknown writing available may be quite limited, consisting of a signature or a few words.

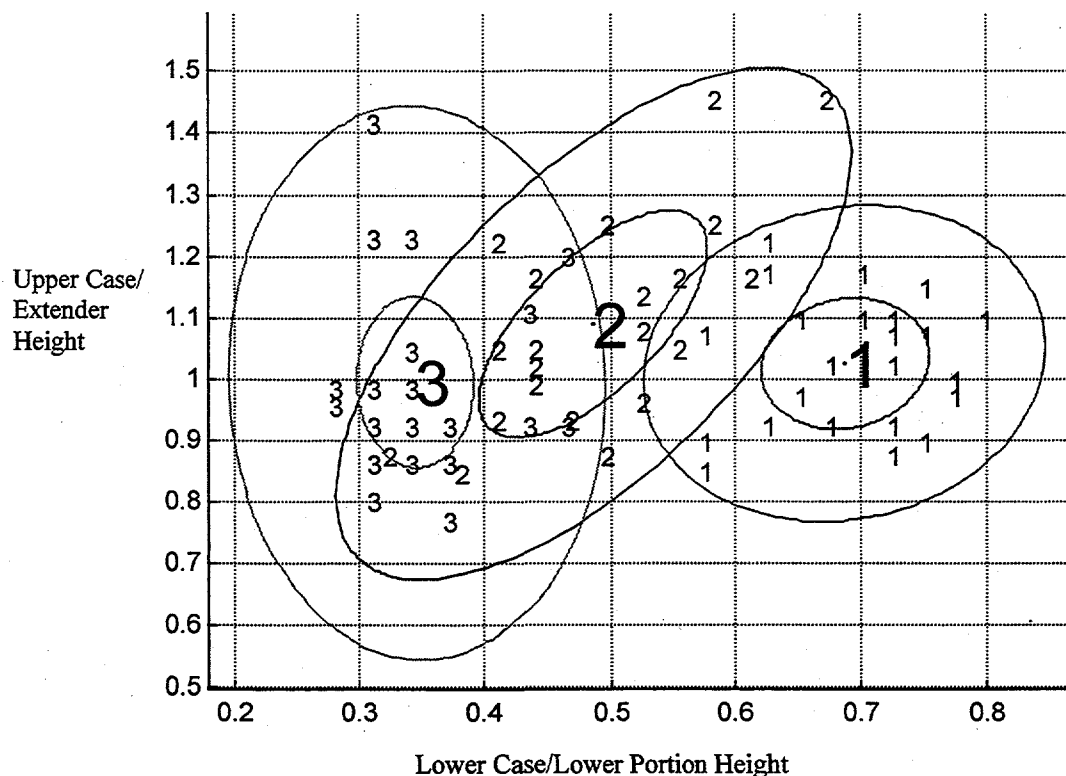


Figure 5. Profile Centroid for Three Exemplars from Primary Study. The y axis represents the upper case/extender height for a given characteristic while the x axis represents the corresponding *a* or lower portion height. The ellipses enclose the extent of the measurements for each exemplar. The extent to which the ellipses overlap corresponds to the degree of similarity in central tendency between exemplars. The smaller ellipses inside each large ellipse represent the uncertainty regions where the different profiles are unique. In particular, each exemplar can be distinguished from the others given measurements falling inside the smaller ellipses.

4.0 Results and Discussion

Table 4 displays results of the cluster analysis performed on the data from the primary study. In the cluster analysis, only the 412 exemplars containing no missing measurements were included. For successively increasing values of α , the number of clusters and number of different size groups generated are given. As seen in the table, the profiles tend to cluster into a couple of very large groups (>50 members), and a number of smaller groups. The number of clusters with more than 20 profiles decreases as α increases, while the number of clusters with 1-5 profiles increases dramatically. Particularly, the number of clusters with one profile, increases from 1 to 83 as α increases from 0.1 to 0.98. This is expected because as α increases, the uncertainty ellipse around each profile becomes smaller, so more profile ellipses have no overlap with any other profiles and therefore become grouped by themselves. Note that since the number of characteristics and amount of data used here is so limited, it is not surprising that for small values of α , few profiles are grouped by themselves.

Table 4. Results of Cluster Analysis of Primary Study Data

α	# of Groups	Number of groups containing n profiles							
		$n=1$	$2 \leq n \leq 5$	$6 \leq n \leq 10$	$11 \leq n \leq 20$	$21 \leq n \leq 30$	$31 \leq n \leq 40$	$41 \leq n \leq 50$	$n > 50$
					0	0	0	0	
0.1	8	1	0	2	2	1	0	0	2
0.2	12	2	2	1	4	1	0	0	2
0.3	15	3	2	2	5	0	1	0	2
0.4	22	2	10	2	3	2	1	0	2
0.5	29	6	13	2	3	1	1	1	2
0.6	34	6	18	1	4	1	1	0	3
0.7	41	6	23	2	4	3	1	0	2
0.8	55	12	31	2	3	4	1	0	2
0.9	80	23	47	4	1	2	0	1	2
0.95	114	33	72	3	2	1	1	1	1
0.98	186	83	100	1	0	1	0	0	1

Figure 6 gives a graphical representation of the clustering for $\alpha=0.2$. In the figure, the data were compressed for two dimensional visualization so that an average of the profile centroids is plotted on the x-axis and an average of the profile variation is plotted on the y-axis. Note that this plot does not contain all clusters and profiles for the case $\alpha=0.2$, but rather provides a visual illustration of the clustering technique. Figure 6 demonstrates the idea of uncertainty and confusion between different groups. In particular, each group lies in a densely populated region around the cluster center, however, the edges of clusters 2, 6, 10, and 11 are not very distinct. Note that one profile from cluster 10 appears to be significantly removed from the rest of the profiles in its group. This is an unfortunate consequence of the loss of information resulting from the compressed centroid and variance, rather than an indication that the profile instead should be grouped in cluster 6. These clusters represent uncertainty regions around the cluster center corresponding to roughly $1-\alpha$. For clusters with a large variation, the clusters are naturally larger in size than clusters with a small variation, because a larger variation results in a larger area within which some percentage of the data lie.

In order to obtain some indication of the effect of adding more characteristics to this analysis, the cluster analysis procedure developed here was also performed on the pilot study data. In this case, only 90 exemplars having no missing measurements were included. The handwriting profiles consist of central tendency and variation for the measurements described in Section 2. The measurements include: height ratios of various letters and the pair 'th', and lateral spacing. Note that for these data, only three measurements are available for some of the characteristics of interest. This is not enough to characterize correlation between characteristics, so only mean and standard deviation was included in this analysis. The results of the cluster analysis are presented in Table 5.

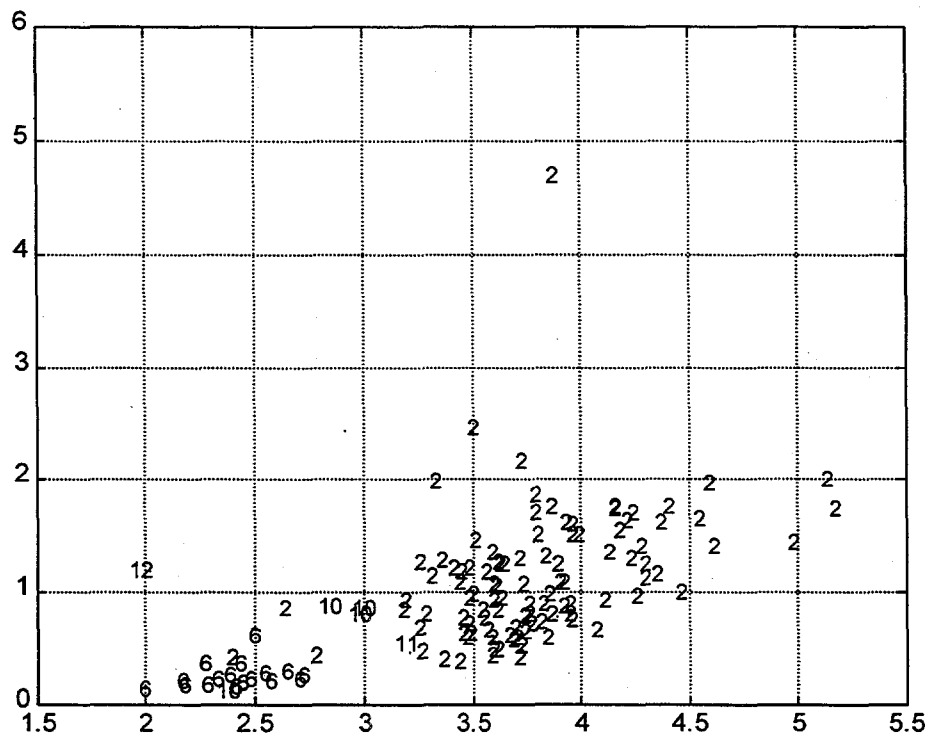


Figure 6. Illustration of Grouping of Handwriting Profiles. Variation is represented on the y axis and centroid on the x axis. Group members are plotted for three groups generated for $\alpha=0.2$. The symbol for each profile represents the group assigned by the cluster analysis. Overlap between groups represents overlap in the uncertainty regions of different cluster.

As seen in Table 5, the group sizes become smaller as α increases, until for $\alpha=0.6$, all profiles are grouped by themselves, indicating that individuals in the pilot study can be differentiated from one another as long as a sufficient number of measurements have been collected. While the number of exemplars available for analysis in the pilot study is limited, the results form Table 5 suggest that with sufficient numbers of measurements in enough habit areas, different individual's handwriting can be differentiated from one another.

Table 5. Results of Cluster Analysis of Pilot Study Data

α	# of Groups	Number of groups containing n profiles				
		$n=1$	2	$3 \leq n \leq 5$	$6 \leq n \leq 10$	$n > 10$
0.1	27	10	12	2	2	1
0.2	55	37	16	2	0	1
0.3	75	60	15	0	0	0
0.5	86	82	4	0	0	0
0.6	90	90	0	0	0	0

5.0 Conclusions

This study is exceedingly limited in the habit areas considered and the samples measured, however, the methods presented here can be used to validate the premises and procedures used in forensic handwriting examination. Using the methods presented here, we can demonstrate the individuality of handwriting and quantify characteristic combinations that are rare and therefore useful in associating or differentiating between an unknown and known source sample. By adding more habit areas to the analysis presented in Section 4, we expect more one or few member groups indicating that many individuals can be differentiated based on the measured characteristics. By performing the analysis presented here on a statistically random sample, we can obtain estimates of the rarity of different characteristic combinations in a population of interest. Expanding the measured habit areas and addressing statistical sampling issues will be the focus of future studies.

6.0 References

Blake, Marty, "Handwriting Individuality: Can We Prove It?", presented to the 46th Annual Conference for the American Society of Questioned Document Examiners, Aurora, CO, 1988.

Hartigan, John A., Clustering Algorithms, John Wiley & Sons, 1975.

Hilton, Ordway, "The Relationship of Mathematical Probability to the Handwriting Identification Problem", *International Journal of Forensic Document Examiners* (1995), v. 1, n. 3, pp 224-229.

Horan, J.J., "A Study of the Frequency of Habits in the Writing of Checks", *Presentation IAFS Meeting, Dusseldorf, 1993*.

Horan, James J., and Horan, George J., "A Study of Numbers", presented at the Tenth International Association of Forensic Sciences Meeting, Oxford, England, 1984.

Horton, Richard A., "A Study of the Occurrence of Certain Handwriting Characteristics in a Random Population", *International Journal of Forensic Document Examiners* (1996), v. 2, n. 2, pp 95-102.

Huber, Roy A., and Headrick, A.M., "Let's Do It by Numbers", presented to the 47th Annual Conference for the American Society of Questioned Document Examiners, Arlington VA, 1989.

Huber, Roy A., "The Uniqueness of Writing", presented to the 48th Annual Conference for the American Society of Questioned Document Examiners, San Jose, CA, August 1990.

McClary, Carl R., "A Study of Baseline Alignment in Signatures and Handwritten Sentences", *International Journal of Forensic Document Examiners* (1997), v. 3, n. 1, pp 35-44.

Moenssens, Andre A., Starrs, James E., Henderson, Carol E., Inbau, Fred E., Scientific Evidence in Civil and Criminal Cases, 4th Ed., Foundation Press, 1995.

Muehlberger, R.J., Newman, K.W., Regent, James, Wichmann, J.G., "A Statistical Examination of Selected Handwriting Characteristics", *Journal of Forensic Sciences*, Vol. 22 No. 1, Jan. 1977, , pp 206-215.

Osborn, A.S., Questioned Document Problems, Boyd Printing Company, 1944.

Rhodes, Henry T.F., "Statistical Approaches to the Identification of Handwriting", *International Journal of Forensic Document Examiners* (1997), v. 3, n. 3, pp. 281-283.

Seber, G.A.F., Multivariate Observations, John Wiley & Sons, 1984.

Starrs, James E., "In the Wake of Harbor Pilots: Document Examiners Given the Twice-Over in Federal Courts", *Scientific Sleuthing Review*, v. 19, n. 2, 1995.

Totty R.N., "The Dependence between Selected Categorical Measured of Cursive Handwriting", *Journal of Forensic Science Society* (1985), v. 25, pp. 217-231.

Zimmerman, Jeannine, "Frequency of Omission, Perseveration and Reversals in Handwriting and Relative Value as Individual Characteristics", *Journal of Forensic Identification* (1990), v. 40, n. 3, pp 127-134.

Appendix A

Exemplar used in Pilot Study

U. S. POSTAL SERVICE
INSPECTION SERVICE

Adam C. Burling	_____	A	_____	T	_____
Catherine E. Dolan	_____	B	_____	U	_____
Edward H. Fallett	_____	C	_____	V	_____
George K. Newburg	Lateral _____	D	_____	W	_____
Mr. Levi Jacobson	_____	E	_____	X	_____
Kitty M. Langdon	_____	F	_____	Y	_____
Margaret P. Hymans	_____	G	_____	Z	_____
Oliver R. Perper	_____	H	_____	bb	_____
Quincy S. Roberts	_____	I	_____	cc	_____
Stancil O. Torque	_____	J	_____	dd	_____
Ulysses T. Velez	_____	K	_____	ff	_____
Warren Bud Smith	Lateral _____	L	_____	gg	_____
Cook F. Young, Jr.	_____	M	_____	kk	_____
Ned Lee Harlow	_____	N	_____	pp	_____
Harold I. Fox, Sr.	_____	O	_____	ss	_____
Don V. Chesterfield	_____	P	_____	tt	_____
1 2 3 4 5 6 7 8 9 0	_____	Q	_____	v	_____
56 West 135th Street	_____	R	_____	x	_____
2033 East Quaker St.	_____	S	_____	yy	_____
9944 Waltham Ave., North	_____				
8877 Hough Avenue, South	_____				

Write below: The above are samples of my handwriting written with my (right/left) hand. I normally write with my (right/left) hand.

Witnessed: Initials _____	Name _____
Date _____	Street _____
	City _____
	Date _____

Appendix B

Exemplar used in Primary Study

HANDWRITING SPECIMEN - PLEASE FILL OUT ENTIRE FORM

Name _____
 Street Address _____
 City _____ State _____
 Place of Birth _____
 Date of Birth _____ Age _____
 Height _____ Weight _____ Sex _____
 Color of Hair _____ Color of Eyes _____
 Right or Left-Handed _____ Build _____
 Occupation/Trade _____
 Present or Last Employment _____
 Name/Address/Relationship _____
 of nearest relative _____

Albert Johnson

Edward Youngberg

Micheal T. Smith

Charles Questa

George C. Kelly

Davies McIntyre

William F. Brown

Raymond G. Taylor

Donald O'Conner

Thomas E. Novak

Peter Fisher, Sr.

Robert Olsen, Jr.

Mary Nan Lamb

Kenneth W. James

Samuel Zimmerman

Elizabeth Vaughn

Franklyn Patrick

Lawrence Harrison

Geoffry Upshaw

one two three four five

six seven eight nine ten

twelve fifteen twenty thirty

forty fifty sixty seventy

eighty ninety hundred and

dollars cash no cents and

1234 N. East Ave. October

5678 S. West Blvd. August

9012 E. North Pl. January

3456 W. South St. December

Your signature Date

Witness Signature Date

Appendix C

Mathematical Discussion of Methods

A handwriting profile consists of the centroid and variation of the following ratios: (1) the capital letter height to the height of the letter "a" within the same word, and (2) extender letter height to the lower portion height for the same letter. Let n be the number of category (1) or (2) ratios in a given exemplar. For each type of characteristic $k=1,2$, the centroid is quantified by the sample mean

$$\hat{\mu}_k = \frac{1}{n} \sum_{1 \leq i \leq n} x_{ik}$$

and the variation is quantified by

$$s_{kj} = \frac{1}{n-1} \sum_{1 \leq i \leq n} (x_{ik} - \hat{\mu}_k)(x_{ij} - \hat{\mu}_j)$$

for $k,j=1,2$ where s_{kj} is the sample variance when $k=j$ and the sample covariance when $k \neq j$. Therefore, the profile for a given exemplar consists of a 2×1 vector \underline{m} of characteristic sample means, and a 2×2 sample variance-covariance matrix S where

$$\underline{m} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} \quad (1)$$

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix}. \quad (2)$$

and

Profiles constructed from (1) and (2) are then grouped using cluster analysis, so that profiles within a group are similar to one another, but profiles between groups are distinct. A modification of the Leader algorithm is used to cluster the profiles. The reader is referred to Hartigan (1975) for a description of the Leader algorithm. The Leader algorithm is a simple and effective method for clustering data. It requires only one iteration through the data and does not require the user to pre-specify the number of groups. The algorithm can be summarized as follows, where cases in this context refer to handwriting profiles.

Specify some distance d_0 which defines the cluster sizes.

1. Begin with case 1 and make it a cluster. Set the center of the first cluster to the first data point and set the number of clusters $M=1$.
2. For cases $i=2,3,\dots,N$,

- a) For clusters $j=1,2,\dots,M$
 Compute the distance d between case i and cluster center j .
 If $d \leq d_0$, assign case i to cluster j , recompute the cluster center as the average of all member profiles, increment i , and return to step 2.
 Otherwise, increment j .
- b) If case i was not assigned to any of the clusters $1 \dots M$, generate a new cluster $M+1$.
 Assign case i to cluster $M+1$, and set the cluster center to case i . Increment the number of clusters M .

We modify the Leader Algorithm by using probabilistic distances rather than absolute distances. In particular, for each profile i and cluster j , we compute a probability that the true mean and variance of profile i are farther from the center j mean and variance than observed from the data, given that the true means and variances are the same. If the probability is greater than some pre-specified value α , then profile i is added to cluster j . Otherwise, profile i is added to some other cluster. This procedure amounts to simultaneous hypothesis tests for equal means and variances of profile i and cluster j , where the rejection of the null hypothesis results in profile i not being placed into cluster j .

Hypothesis tests for the mean and variance of a multivariate population have been studied extensively. In this research, the test for equality of variance of profile i and cluster j is performed first using a standard likelihood ratio test assuming unequal mean vectors (see Seber (1984), p. 102). If the test for the variance results in acceptance of the null hypothesis, then the sample variances for profile i and cluster j are pooled, and Hotelling's T^2 -test for equal means is performed.

Suppose, in truth, that profile i has the same mean and variance as cluster j center. Then the probability that profile i will correctly be placed in cluster j is the probability that the hypothesis tests for both mean and variance accept the null hypothesis. Since the sample mean and variance of a multivariate population are independent, we have

$$\begin{aligned} 1-\alpha &= P\{\text{mean test and variance test accept } H_0 \mid H_0\} \\ &= P\{\text{mean test accepts } H_0 \mid H_0\} P\{\text{variance test accepts } H_0 \mid H_0\} \\ &= (1-p_0)^2 \end{aligned}$$

where p_0 is the significance level of each individual test. Therefore, to ensure that a profile i will be placed in the appropriate cluster j with probability $1-\alpha$, we set the significance level of each hypothesis test to $p_0 = 1-(1-\alpha)^{1/2}$.

From this we can now formalize the modified Leader algorithm for clustering handwriting profiles as follows.

Determine desired error probability α . Set $p_0 = 1-(1-\alpha)^{1/2}$.

- 1) Begin with profile 1 and make it a cluster. Set the center of the first cluster to the sample mean and variance of profile 1. Set the number of clusters $M=1$.
- 2) For profiles $i=2,3,\dots,N$,
 - a) For clusters $j=1,2,\dots,M$
 - i) Perform a hypothesis test for equality of variance of profile i and cluster j . Compute the significance level p_V .
 - ii) If $p_V < p_0$, pool the variance of profile i and cluster j , perform a hypothesis test for equality of mean of profile i and cluster j , and compute the significance level p_M . Otherwise, increment j and return to step 2.a.

- iii) If $p_M < p_0$, pool the means of profile i and cluster j . Assign profile i to cluster j and set the center of cluster j to the pooled sample mean and variance. Otherwise, increment j and return to step 2.a.
- iv) Increment i and return to step 2.
- b) If profile i was not assigned to any of the clusters $1 \dots M$, generate a new cluster $M+1$. Assign profile i to cluster $M+1$, and set the center to the sample mean and variance of profile i . Increment the number of clusters M . Increment i and return to step 2.

As described here, this algorithm makes one pass through the data where each profile is placed in the appropriate cluster. It is possible that as the algorithm progresses, the center of a given cluster moves around enough so that profiles initially assigned to it no longer belong to the cluster in terms of the criteria used in this algorithm. Therefore, a second stage to this algorithm is implemented in practice where each profile is tested for equality of mean and variance with its corresponding cluster center. Profiles for which either hypothesis test rejects the null hypothesis are reassigned according to the above procedure. While this does occur, this seems to be a relatively small issue since typically less than 5% of profiles become reassigned on the second pass through the data.

The value selected for α directly affects the number and size of the clusters generated, but it also affects the probability that a given case will not be placed in the proper cluster. In particular, we can think of a cluster as forming $1-\alpha$ level confidence regions around the cluster center mean and variance. A larger value for α reduces the confidence level and hence the size of the region around the cluster center, resulting in smaller cluster sizes. However, as α increases, so does the probability that a profile matching a given cluster in terms of mean and variance will fall outside the confidence region for that cluster. In terms of associating an unknown source exemplar to a known source exemplar, this would correspond to an increased probability of not associating two profiles written by the same individual.

Validation of Modified Leader Algorithm using Simulated Data

A simulation study was performed to gain some understanding of the strengths and weaknesses of the algorithm used here to group handwriting profiles. The study was performed as follows. The centers and sizes of the groups generated through cluster analysis of the handwriting data were taken as truth. The profiles are obtained by selecting a truth group at random according to probabilities obtained from the relative sizes of each group. Once the truth group is selected, 10 measurements from a multivariate normal distribution containing the truth group mean and dispersion matrix are generated, and a profile constructed from the sample mean and variance/covariance matrix of the simulated observations. A large number of profiles are simulated and the cluster analysis performed.

The results of the cluster analysis are compared to the truth data for effectiveness of clustering the same truth group profiles. If the initial groups of data are well separated, we would ideally expect the fraction of incorrectly grouped profiles to be close to α , since it is the probability the hypothesis tests will incorrectly reject a profile. If the initial truth groups contain significant overlap, then the chances are increased that a given profile will not be grouped into the correct group, but will instead be placed in a nearby cluster. In this way, the algorithm presented here will be affected by both type I error (probability of a group incorrectly rejecting a member profile) and type II error (probability of a group incorrectly accepting a member profile). Another factor complicates our ability to predict the performance of the cluster analysis algorithm. Namely, as the clustering algorithm proceeds, the group centers are allowed to move

around as profiles are added to and deleted from groups. As a result, groups could drift far away from other groups increasing the separation or they could drift into other groups increasing the overlap.

The results of a simulation for $\alpha=0.2$ is given in Table 5. In the table, the number of profiles simulated, the estimated fraction grouped incorrectly, and the standard error (s.e.) of the fraction grouped incorrectly are given. In the first simulation, the truth groups correspond to the centers of the groups constructed during the cluster analysis of the primary study data presented in Section 4. The centers of all the truth groups used are plotted in Figure 7. In the second simulation, the group centers are obtained in the same manner, except only half of the truth groups are used for truth groups. Truth groups omitted in the second simulation are represented by "***" in Table 5 where groups included contain the simulation results.

The combined effects of type I and type II error can be clearly seen in Table 5. In particular, in the first simulation, some of the truth groups are formed very effectively, while some have an error rate roughly twice the values $\alpha=0.2$. All of the groups whose error rate is much greater than α (groups 2,3,4, 8 and 10) are in the center of the plot contained in Figure 7, indicating that the profiles in these clusters are in the middle of all the data. Therefore, one would expect these groups to be more poorly separated than groups on the edges of the data in some sense. In the second simulation, all the odd numbered truth groups were removed. Doing this removes several groups which are very close to others. In this case, we would expect the type II error to be much smaller in this case and the clustering to perform much more effectively. This is evident by the fraction of incorrect groupings appearing on the right side of Table 5. Improved performance as the separation in truth groups increases is a common property of cluster analysis algorithms.

Table 6. Results of Simulated Grouping of Handwriting Profiles for $\alpha=0.2$

Group ID	Number of Profiles Simulated	Est. Fraction Grouped Incorrectly	S.E. of Fraction Grouped Incorrectly	Number of Profiles Simulated	Est. Fraction Grouped Incorrectly	S.E. of Fraction Grouped Incorrectly
1	23604	0.033	0.0012	28537	0.016	0.00074
2	11707	0.398	0.0045	**	**	**
3	2071	0.408	0.0108	2414	0.234	0.00861
4	1654	0.407	0.0121	**	**	**
5	3006	0.271	0.0081	3652	0.063	0.00403
6	1934	0.145	0.0080	**	**	**
7	1586	0.261	0.0110	2037	0.239	0.00944
8	806	0.366	0.0170	**	**	**
9	230	0.174	0.0250	309	0.236	0.0242
10	579	0.402	0.0204	**	**	**
11	101	0.040	0.0194	131	0.053	0.0197
12	102	0.127	0.0330	**	**	**

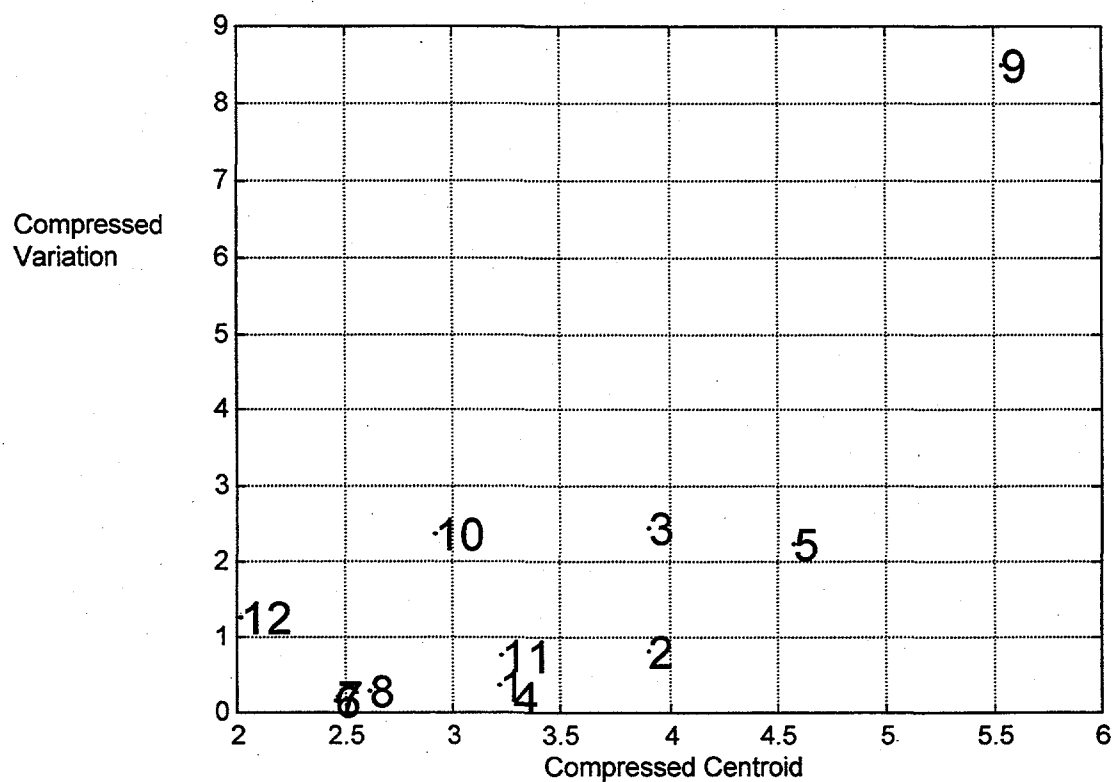


Figure 7. Truth Group Centers used in Handwriting Profile Cluster Validation. Each number represents a truth cluster used in the simulation study to validate the cluster analysis algorithm. Note that in order to represent the cluster centers in two dimensions, the x axis plots a weighted average of the centroids for the cluster center, and y axis plots a weighted average of the variance for the cluster center.

Distribution

No. Of Copies External

1 Todd Anderson
 Office of Special Technology
 10530 Riverview Rd., Bldg. 3
 Fort Washington, MD 20744

2 Patricia Manzollilo
 USPISFL
 225 N. Humphreys Blvd.
 Memphis, TN 38161-0003

1 Carrie Whitcomb
 Manager, Forensic Science Services
 USPIS Crime Laboratory
 22433 Randolph Drive
 Dulles, VA 20104-1000

No. Of Copies Internal

26 Pacific Northwest Laboratory

 RG Clemmer (3) K6-48
 LE Glaspell K8-03
 RC Hanlen (3) K6-48
 KH Jarman (10) K5-12
 SW Martin (2) K8-21
 Information Release (7) K1-06