Measuring the Frequency Occurrence of Handwriting and Hand-Printing Characteristics

NIJ Award 2010-DN-BX-K273

Mark E. Johnson
Dept. of Statistics, University of Central Florida
mejohnso@mail.ucf.edu

Thomas W. Vastrick
Forensic Document Examiner
vastrick@yahoo.com

Michèle Boulanger
Dept. of International Business, Rollins College
mboulanger@rollins.com

Ellen Schuetzner
Forensic Document Examiner
ejsqde@sbcglobal.net

Measuring the Frequency Occurrence of Handwriting and Hand-Printing Characteristics

Abstract

This report describes the results from a National Institute of Justice funded statistical research project through the National Center of Forensic Science at the University of Central Florida. The motivation of the study was to strengthen the statistical basis for handwriting comparisons, following the recognition that the discipline of forensic document examination was facing increasing judicial scrutiny under the Daubert guidelines as recognized by the profession and subsequently reported in the National Research Council report, *Strengthening Forensic Science in the United States: A Path Forward* (2009). In response, this project's objectives were to develop statistically valid frequency occurrence proportions for selected characteristics of handwriting and hand printing based on specimen samples representative of the United States population, to provide practitioners of forensic document examination with a statistical basis for reliability and measurement validity and to provide courts with the requested supporting data

The project produced an initial set of over 2500 precise handwriting and hand printing features that were subsequently reduced to 903 features which passed an attribute agreement analysis and to 786 that were utilized in this project. These attribute features (presence/absence) can be unambiguously identified by forensic document examiners. Handwriting samples from over 1500 writers were collected representing a broad spectrum of contributors intended to be representative of the US adult population. Meeting the pre-specified population representation led to the selection of a subset of 880 cursive specimens and 839 hand printed specimens that closely approximated the demographic proportions represented in the US. The analysis of these specimens yielded numerous specific frequency occurrence proportions. Additional analyses have shown quantitatively the extent to which demographic features such as age, gender, ethnicity, education, location of second/third grade training and handedness impact the presence/absence of features. An immediate benefit of the databases analysis has been a detailed assessment of the scope of the appropriateness of the product rule.

This project relied heavily on international standards and appropriate statistical methodology to develop the sampling protocols.

Table of Contents

Measuring the Frequency Occurrence of Handwriting and Hand-Printing Characteristics

Executive Summary

Forensic evidence has come under increasing scrutiny in the past several years. A pattern has emerged within the specialty discipline of forensic document examination involving admissibility rulings in which judges were admitting forensic document examination as reliable under the Daubert guidelines but chastising the discipline for having limited empirical bases. At a Questioned Documents Symposium in Ames, Iowa in 2008, investigators Vastrick and Schuetzner along with Forensic Document Examiners Kathleen Storer and Karen Runyon met and developed an outline of what was to eventually become this research project.

Publication of the National Research Council report, *Strengthening Forensic Science in the United States: A Path Forward* (2009) further established that forensic science in general needed additional scientific underpinnings. Included in the report was a section on handwriting comparisons for which these limitations were identified. In particular, this report noted with respect to handwriting analysis the following:

> "The scientific basis for handwriting comparisons needs to be strengthened. Recent studies (Kam et al, 1997) have increased our understanding of the individuality and consistency of handwriting and computer studies (Srihari et al, 2002) suggest that there may be a scientific basis for handwriting comparison, at least in the absence of intentional obfuscation or forgery. Although there has been only limited research to quantify the reliability and replicability of the practices used by trained document examiners, the committee agrees that **t**here may be some value in handwriting analysis."

Recognizing deficiencies in empirical bases, the National Institute of Justice launched several research initiatives to address these issues. Based on the groundwork performed prior to 2009, The National Center for Forensic Science at the University of Central Florida received a $550,000.00 grant to establish frequency of occurrence proportions for specific cursive and hand printed characteristics by writers of diverse demographic backgrounds representative of the U.S. population. The databases are designed for use by research investigators, forensic document examiners in their daily work, and forensic document examiners in courtroom situations during which questions of empirical bases are raised. The potential applications of the research and its results described in this report include use in both civil and criminal cases. The objectives of the project in the original proposal were, as follows:

1. Develop statistically valid frequency occurrence proportions of handwriting and hand printing characteristics based on specimen samples from throughout the United States;
2. Provide practitioners of forensic document examination with statistical basis for reliability and measurement validity to accurately state their conclusions and assess complexity; and
3. Provide courts with the reliable data needed to understand the underlying statistical basis for the conclusions.

A significant number of handwriting and hand printing characteristics were selected based on the belief that these characteristics were objective in nature – avoiding terms like "long stroke" or "sharp curve" that are susceptible to subjective interpretations from different classifiers. Approximately 2500 characteristics were initially selected but were eventually pared down to 903 characteristics via an Attribute Agreement Analysis (AAA). A multi-level sampling approach was utilized to gather handwriting and hand printing specimens. Over 1500 writers provided specimens to the study. The primary goal of this project is to determine how many of the writers did or did not utilize the selected handwriting and hand printing characteristics in their writing specimens. Using standard statistical sampling methodologies, frequency occurrence proportions for each characteristic within the general population can be determined within an established error rate for the sampling methodology. The utility of this database is demonstrated in a results section that addresses the validation of the concept of degree of individuality to handwriting and the application of the product rule in handwriting assessments among other statistical evaluations.

Throughout the project, strict adherence to standards has been paramount. This project is a statistical study, not a forensic study. As such, the methodologies used are driven by standards and best practices from within the statistical field. The resulting data from this project has the potential for significant effects on the profession of forensic document examination. One of the basic axioms of handwriting comparison is that no two writers utilize the exact same set of handwriting characteristics. The database created by this research and the resulting frequency occurrence proportions provides the forensic and judicial communities some empirical data concerning actual statistics with which to assess this axiom.

Forensic document examination standards benefit by strengthening their foundations through the data from this project. For example examiners can utilize the project data as a part of their estimation of confidence designated by the NRC Report (2009) in stating conclusions for forensic document examinations as currently described in ASTM-E1658 (2008), which was also specifically referenced and quoted in NRC Report (2009) and now codified through SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners (2012). Although the opinions in brief coupled with their elaborations provide a reasonable estimate of confidence, some quantification of these levels would be welcome to the forensic and judicial communities. The frequency occurrence proportions

developed in this project potentially can serve to provide these categories with supporting calculations.

The scientific literature describing various sources that affect handwriting is conveniently summarized in the compendium by Huber and Headrick (1999). This reference captures the state of the art as of 1999 and suggested demographic and other factors to be considered and note both extrinsic and intrinsic factors that influence handwriting. Ultimately to address Huber and Headrick's summarization, this study focused on the demographic factors of age, gender, ethnicity, education, location of second/third grade schooling and handedness while controlling for other factors such as temporal state and geographic location. The original set of specimens collected from a wide spectrum of participants provided an initial large collection which was eventually pared down to achieve what is considered to be a representative sample from the entire target population.

This study is not the first on the subject of frequency occurrence in handwriting. The full report summarizes the associated literature. The specific approach in collecting samples that match the US demographics provides unique insight and substantive data into the quantitative relationship between the presence/absence of readily identifiable features and key demographic factors as noted by Huber and Headrick.

In short the basic premise of this project is simple; 1) collect handwriting specimens from writers selected to constitute a "representative" sample, 2) examine each and note the presence/absence of predetermined handwriting characteristics, and 3) calculate the ratio of characteristic presence and characteristic absence for each specimen. However, the complexity of the selection process of writers and the magnitude of the characterization effort of their writing, along with the pilot testing necessary for each step of the methodology provided a challenging effort to the research team, particularly keeping in mind the required use of standard methodologies and best practices.

Absence of a Sampling Frame and Acquisition of Samples. In many sampling applications, a sampling frame is available which consists of all units in the relevant population. In this project a sampling frame was not available. No frame of all possible individuals from which a random sample could be drawn exists available. The target population for this study was defined as adults 18 years old or older who are capable of providing writing samples by hand and in English (preferably both cursive and hand printed forms). Younger writers are increasingly less trained in cursive writing, so some specimens provided are strictly the hand printed version.

The rationale for our approach was described by Boulanger, Johnson and Vastrick (2013), as follows:

> "Thus our approach to data collection changed from a probabilistic sampling process to the development of a data collection process that will lead to a large sample of "writers" deemed representative of the target population. The approach

we followed was based on a study done to evaluate the performance of the national telecommunications network before the breakout of the monopoly service provider, AT&T (Boulanger Carey et al, 1999; ASTM-E105-04, 2004; ASTME141-91, 1991). There, as in our situation with handwriting, it was not possible to construct a sampling frame of all the potential telecommunication paths in the US and a multi-level sampling approach based on identification of strata and clusters was developed to lead to a quasi-representative sample."

The planning approach used had seven steps governing the data collection process was as follows:

1. Research factors influencing handwriting
2. Define stratification variables based on key factors influencing handwriting and key variables describing the target population
3. Define strata for selected variables
4. Estimate proportions within the strata of the target population
5. Define the data collection process for obtaining a representative sample of the target population, deemed so by meeting the quota guidelines
6. Provide guidelines for the data collection process to the data collectors
7. Audit the data collection process for adherence to the plan and for quality control

In the absence of a sampling frame, the investigators opted to obtain samples from a collection of adult participants who are reasonably representative of the target population regarding demographics and other factors known to have influence. In order to obtain the set of specimens corresponding to a representative sample of writers, a large set of specimens were obtained and then at the analysis stage, a subset of the population enjoying common overall demographic characteristics to the total population was determined for the ultimate determination of individual and multi-dimensional frequencies. The paring process was conducted by the statisticians using appropriate methodologies to achieve the demographic thresholds as closely as possible

A large number of forensic document examiners and others were called upon to collect specimens. In an effort to prevent unintentional variations in the entire collection process all handwriting specimen forms were printed at one time from one source. In addition, a bulk supply of non-retractable ballpoint pens were purchased and distributed for use in specimen collection. Project management worked with the UCF Institutional Review Board to assure compliance with federal regulations involving human subject participation. In particular, contributors who decided to withdraw their participation and specimens part way through the effort were free to do so. The participants providing specimens were volunteers who responded to the societal benefit by contributing to a research project at the university. No payments were made to any participants providing specimens.

The specimen collectors were provided general guidance as to the categories of participants in the specimen collections. The intent was to meet a fixed minimal quota specification of 80% of the sample for each key stratification variable. Since

we were able to come close to achieving the pre-determined strata ranges, the specimen collectors were able to accomplish their ultimate goal through this random-based stratification collection process.

In summary, great lengths were taken to ensure that the samples collected would satisfy the representativeness of our sample with respect to the target population while at the same time minimizing any negative effects of pure quota or pure random data collection.

Pilot Studies. In terms of individual characteristic delineation, a major initial step of this project was to identify and define multiple individual characteristics of letters (cursive and hand printed), numbers and symbols, then prepare a database for classifiers to check the presence or absence of these features. Over 2500 individual characteristics were defined initially. An Attribute Agreement Analysis (AAA) was designed and implemented to test each characteristic for objectivity and reproducibility of results. The AAA also simultaneously tested the method of presentation (the database) for the same objectivity and reproducibility of results. The statistical aspects of the attribute agreement analysis approach used is embodied in ISO TR 14468 and supported by the international statistical community. A total of 903 characteristics survived the attribute agreement analysis. The surviving characteristics had one-hundred percent agreement by multiple examiners across multiple handwriting specimens, including replicated reviews by the same examiner. A single disagreement was grounds for dismissal as this instance would shed doubt on the reliability of the data. Given the large number of features, a simple presence/absence response was in order for the hundreds of features inspected. A benefit of the attribute agreement analysis was the elimination of characteristics that generated discrepant opinions and to produce a list of features worthy of assessment over the 1500 specimens collected.

The design of the attribute agreement analysis was to have three ABFDE-certified forensic document examiners classify five complete specimens (both cursive and printed) including two replicates for each of the original 2500 characteristics. The original set of over 2500 features was reduced to 903 (485 for cursive and 418 for printed specimens) following this analysis. Some letters were fully excluded from the project as were all numbers and punctuation. Further reductions were also made subsequent to the AAA to reach the final project numbers of 435 for cursive and 351 for printed specimens.

A digital method for managing the specimen handwriting for classification was initially tested using AAA at the 69th Annual Meeting of the American Society of Questioned Document Examiners. Based on the results of the attribute agreement analysis pilot study, investigators Boulanger, Johnson and Vastrick collectively decided that the devised process was not sufficiently reliable for this project. As a result it was concluded that classifiers would not use this particular or any handwriting management system, instead utilizing the original handwriting specimens for classification. An independent illustration was devised that assisted classifiers by denoting the location of each specific character throughout the text of

the handwriting form. This process was subjected to the AAA study described above that led to the 903 features analyzed in the project study.

The handwriting specimens were collected and submitted to the National Center for Forensic Science (NCFS). Each specimen was provided a writer number which was subsequently used for any computer referencing. The biographical data and other extrinsic/intrinsic collected factors were recorded and each form was placed in a document protection sheet. The NCFS then distributed the specimens to classifiers for classification and data entry into the database. Upon completion of the analysis the forms were returned to the NCFS where they are currently being stored per government privacy requirements.

The database was developed using a common commercially-available database software program modified specifically for this project. Forensic document examiners provided a combination of characteristic descriptions with accompanying example images that the database expert, Heather Burske, incorporated into the user-friendly database. Upon completion of the classification process, Burske submitted a completed spreadsheet simultaneously to investigators Boulanger, Johnson and Vastrick. The resulting spreadsheet required considerable data preparation prior to the analysis stage. The data preparation work frequently can occupy 80-90% of the labor of the analysis stage, and this project was no exception. Appendix 3 of the full report details the lengths taken to produce viable data for analysis. Careful attention to the coding of the collected data and identifying correctable problems provides confidence in the ultimate conclusions.

Demographics. In examining the demographics associated with the 1517 specimens, we recognized that some obvious selections could be made to attempt to achieve the prescribed quotas. These decisions yielded demographics for the set of 880 cursive specimens ultimately analyzed summarized in Tables 1 through 6. Aside from the middle age category (30-50 years coded as 40) and the Hispanic ethnicity, the quotas were met.

Age	count	%	Quota %
24	332	38.2%	> 20%
40	243	27.9%	> 30%
60	295	33.9%	> 30%

Table 1 – Cursive Population Sampling Based on Age

Gender	count	%	Quota %
Female	390	44.6%	> 40%
Male	484	55.4%	> 40%

Table 2 – Cursive Population Sampling Based on Gender

Ethnicity	count	%	Quota %
African-American	95	11.4%	> 10%
Asian	49	5.9%	> 4%
Caucasian	612	73.4%	> 55%
Hispanic	83	10.0%	>11%

Table 3 – Cursive Population Sampling Based on Ethnicity

Right or Left	count	%	Quota %
Left	78	9.0%	> 5%
Right	791	91.0%	> 75%

Table 4 – Cursive Population Sampling Based on Handedness

Education Level	count	%	Quota %
High School or less	279	32.4%	> 30%
HS plus	582	67.6%	> 50%

Table 5 – Cursive Population Sampling Based on Level of Education

$2^{nd}/3^{rd}$ Grade Education Location	Count	%	Quota %
US	626	89.9%	> 70%
Not US	70	10.1%	> 10%

Table 6 – Cursive Population Sampling Based on Location of Early Education. Forty-seven of the fifty states were represented in the sample collection.

Subsequent results in this report are based on analyses of the 880 cursive specimens, collectively designated the "cursive project sample".

A key deliverable in this project is an estimation of the frequencies of the presence/absence of features within the cursive project sample. There were 435 specific characteristics examined for the cursive specimens, covering 49 lower case and upper case letters. The lower case letters "a", "b" and "d" did not have any features that survived the attribute agreement analysis described previously. The entire set of proportions of presence of the features is given in the full report. For the cursive writing, those features that passed the AAA were tallied with respect to presence/absence of the features. The frequency for each feature across the overall 880 specimens was determined. For example, for cursive lower case "c" (CLCC) there were 865 specimens examined for lower case "c" features (15 of the 880 cursive specimens did not have a qualifying presence of the character), yielding observed proportions of presence, as follows:

CLCC 2	0.94451
CLCC 3	0.88208
CLCC 4	0.22197
CLCC 5	0.86012
CLCC 6	0.10058

Table 7 – Examples of Frequency Occurrence Proportions

The standard deviation of these proportions is less than 0.017 in general and further reduced for proportions close to either 0 or 1.

Huber and Headrick (1999) describe *qualitatively* various demographic features that influence handwriting in general. With the cursive project sample established and in conjunction with the associated demographics, we can *quantitatively* assess character features as a function of age, gender, ethnicity, education level, location of cursive training, and handedness. For each combination of character feature (435 for cursive) and demographic (6 possibilities), the association as measured by Fisher's exact test has been run. A significant association implies that the demographic variable influences the presence/absence of a feature. Table 8 summarizes the results for all 435 cursive features across the 6 demographic variables.

p-value range	Age	Location 2nd or 3rd Grade	Gender	Education	Ethnic	Right or Left
<.0001	13.1%	11.9%	2.5%	1.4%	0.2%	0.2%
<.001	17.0%	19.0%	8.7%	3.9%	1.1%	0.2%
<.01	28.2%	26.8%	16.1%	11.2%	6.9%	1.1%
<.05	40.8%	35.3%	28.7%	21.8%	16.3%	5.3%
<.1	50.0%	42.2%	37.6%	28.4%	23.4%	9.4%
>.95	1.8%	12.8%	6.9%	7.1%	3.2%	18.1%
1	1.4%	12.8%	6.9%	7.1%	1.6%	18.1%
Demographic Row Total Basis	870	696	874	861	834	869

Table 8 - Percentage of 435 Cursive Features Having Indicated p-value Range.

Table 8 is arranged from left to right according to the strength of association (stronger to weaker). Overall, age has the greatest bearing on the number of features present with Location of second/third grade training a close second. Over a quarter of the features considered exhibit an effect on the presence/absence due to age of provider or location. Gender and education also exhibit a significant signal for many of the features (many more than would be expected due to chance alone). The ethnic category (restricted to Caucasian, African American, Hispanic and Asian) also influences a number of features presence/absence. Only handedness appears

to have little to do with influencing presence/absence with percentages matching those that would be observed due to chance alone.

Data preparation for the hand printed specimens followed the same process as for cursive specimens. Data preparation germane to the hand printed specimens is described in Appendix 4. There were 1515 specimens in the original spread sheet which was subsequently reduced to 839 specimens. The resultant demographics for the 839 are provided in the Tables 9 - 14. As can be seen, the quotas are nearly all met.

Age	count	%	Quota %
24	321	38.8%	> 20%
40	231	27.9%	> 30%
60	276	33.3%	> 30%
Unknown	11		

Table 9 – Hand Printed Population Sampling Based on Age

Gender	count	%	Quota %
Female	336	40.3%	> 40%
Male	497	59.7%	> 40%
Unknown	6		

Table 10 – Hand Printed Population Sampling Based on Gender

Ethnicity	count	%	Quota %
African-American	94	11.5%	> 10%
Asian	48	5.9%	> 4%
Caucasian	572	70.1%	> 55%
Hispanic	87	10.7%	>11%
Mixed	3	0.4%	
Native American	3	0.4%	
Other	4	0.5%	
South Pacific	5	0.6%	
Unknown	23		

Table 11 – Hand Printed Population Sampling Based on Ethnicity

Right or Left	count	%	Quota %
Left	71	8.6%	> 5%
Right	757	91.4%	> 75%
Ambidextrous	4		
Unknown	7		

Table 12 – Hand Printed Population Sampling Based on Handedness

Education Level	count	%	Quota %
High School or less	282	34.3%	> 30%
HS plus	541	65.7%	> 50%
Unknown	16		

Table 13 – Hand Printed Population Sampling Based on Level of Education

2nd/3rd Grade Education Location	Count	%	Quota %
US	637	90.4%	> 70%
Not US	68	9.6%	> 10%
Unknown	134		

Table 14 - Cursive Population Sampling Based on Location of Early Education. Forty-seven of the fifty states were represented in the sample collection.

The subsequent results in this report are based on analyses of the 839 hand printed specimens, collectively designated the "hand printed project sample".

A key deliverable in this project is an estimation of the frequencies of the presence/absence of features within the hand printed project sample. There were 351 specific characteristics examined for the hand printed specimens, covering 50 lower case and upper case letters. The lower case letters "c" and "i" did not have any features that survived the attribute agreement analysis described previously. The entire set of proportions of presence of the features is given in the full report. For the hand printing, those features that passed the Attribute Agreement Analysis were tallied with respect to presence/absence of the features. The frequency for each feature across the overall 839 specimens was determined.

As noted earlier in this report, Huber and Headrick (1999) describe *qualitatively* various demographic features that influence handwriting in general. With the hand printed project sample established and in conjunction with the associated demographics, a *quantitative* assessment of character features as a function of age, gender, ethnicity, education level, location of cursive training, and handedness has been established. For each combination of character feature (361 for hand printed) and demographic (6 possibilities), the association as measured by Fisher's exact test has been run. A significant association implies that the demographic variable influences the presence/absence of a feature. Table 15 summarizes the results for all 351 features across the 6 demographic variables. Note that the demographic row total basis values vary depending on the number of unknowns for the demographic category, a sub-category deliberately not included (e.g., Native American for ethnic or ambidextrous for right or left which have very small sample sizes), or some specimens not examined for a given letter and demographic (hence, the range of row totals provided).

p-value range	Age	Location 2nd or 3rd Grade	Gender	Education	Ethnic	Right or Left
<.0001	11.9%	0.3%	5.0%	1.4%	6.4%	0.6%
<.001	16.3%	0.8%	9.1%	4.2%	6.9%	1.1%
<.01	24.9%	1.9%	18.0%	13.0%	10.0%	1.4%
<.05	34.9%	3.9%	24.1%	21.1%	18.8%	5.0%
<.1	41.8%	11.6%	31.3%	27.1%	25.2%	9.4%
>.95	3.3%	27.4%	17.7%	2.2%	10.0%	30.7%
1	2.8%	27.4%	17.7%	2.2%	6.9%	30.7%
Demographic Row Total Range	606-810	501-689	610-815	602-805	587-784	607-810

Table 15 - Percentage of 351 Hand Printed Features Having Indicated p-value Range.

Table 15 is arranged from left to right according to the strength of association (stronger to weaker) that was observed with the cursive specimens. The results for the location of 2nd/3rd grade education for hand printing differ greatly from the corresponding results for cursive. Overall, age has the greatest bearing on the number of features. Over a quarter of the features considered exhibit an effect on the presence/absence due to age of the writer. Gender and education also exhibit a significant signal for many of the features. The ethnic category (restricted to Caucasian, African American, Hispanic and Asian) also influences a number of features presence/absence. Handedness continues to have little to do with influencing presence/absence with percentages matching those that would be observed due to chance alone.

Product Rule Analyses. The product rule is recognized as a convenient tool if in fact it applies. The datasets produced in this project provides numerous instances for testing the appropriateness of the product rule with respect to presence or absence of combinations of characteristics. Since there are 436 distinct cursive features available for analysis then, there are a total of 94,830 possible pairs of cursive features that could be considered. There are 97.01% of all cursive feature pairs that have correlations in the range (no more than plus or minus 0.2) for which the product rule is satisfactory. Frequently, the large correlations occur for two features on the same letter. Since there were 361 distinct hand printed features available for analysis then, there are a total of 64,980 possible pairs of hand printed features that could be considered, 57,862 of which had a non-missing coefficient of correlation (due to no variability seen in at least one of the features in the pair). There is 98.55% of the 57,862 combinations of hand printed features that have a

coefficient of correlation between -0.2 and +0.2 for which the product rule is satisfactory.

The frequency occurrence proportions using the project specimens can be used for numerous other investigations. This study has been designed in such a manner that it can be enhanced through further projects that increase the population of specimens or features.

The conclusions in the form of the frequency occurrence proportions of cursive and hand printing characteristics based on the collected specimens are found in Section 4 of the full report. Forensic document examiners testifying in court now possess a product from which frequency occurrence data can be offered, providing courts with the answers they have been requesting for years. The profession of forensic document examination will need to review the provided data and assess the various ways in which this material can be incorporated into daily examination use. For example, an examiner could query features found within what appears to be a generic form of limited handwriting in order to assist them in the determination as to whether the entry is sufficiently complex or individualistic for comparison purposes. In addition, the project data could provide quantitative assistance in estimating the confidence of conclusions in our conclusions scale. These and other potential uses must be suggested, discussed, and tested before gaining a profession-wide consensus for the adoption of examination methodology uses. It is not for any one person to make these decisions. That said, the data is now available and those discussions can begin.

There is a high potential for misuse of the information in this project. As such many cautionary comments are warranted.

It should be understood that the scope of characteristics examined by forensic document examiners in the course of any examination will far exceed the numbers presented in this project by many factors. This project has just scratched the surface of the detail that is reviewed and is designed to give the user an appreciation of the probabilistic level of individuality in handwriting. Forensic document examiners should not be limited solely to the features listed in this project as doing so would be a specific misuse of the intent of this project and the scope of examinations.

One should not apply any inverse application to the frequency occurrence proportions. If the presence of a characteristic has a frequency occurrence proportion of 0.25, it cannot be assumed that the absence of this characteristic has a frequency occurrence of 0.75. The reason for this is variation in handwriting. This study not only applies a present/not present format for establishment of characteristic frequency but also applies presence priority. Per the example if the character being reviewed was present once but absent one hundred times within the handwriting specimen, the database box would reflect the presence of the characteristic. Likewise if the feature at issue was whether a specific characteristic

was not present then one could not apply the inverse of the frequency occurrence proportion if the feature was present for the same reason.

This project in no way promotes or describes methodology for the comparative examination of handwriting based solely on the results of this research. Should an individual attempt to simulate or trace another's writing, it would be expected to find a significant degree of characteristic agreement in the features described in this report. However, the vital features of line quality, blunt ending strokes, hesitations, pen lifts, and other features of simulations or tracings (that are used by qualified forensic document examiners to expose such activity) are not an aspect of this study.

There are not necessarily homogenous reasons for the notation that a characteristic is present or absent (Figure 1). As such, the presence or absence of any characteristic merely begins to illustrate the level of differences in handwriting characteristics and provides an appreciation for the level of uniqueness to any given characteristic. The reader should understand that there are other factors that provide additional contributions to the determined level of heterogeneity of handwriting for instance, the different reasons for which a box was checked or not checked in the database.

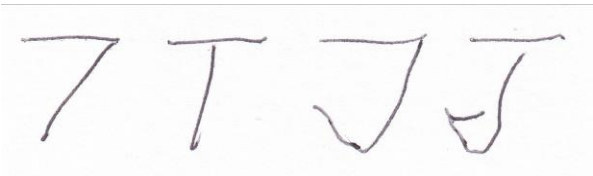

Figure 1 – Cursive Upper Case T (CUCT) Feature 14 "Cap is approximately straight" applies to each of the above versions of a cursive upper case "T". This figure illustrates the non-homogenous reasons that boxes are checked and why presence/absence is a small aspect of individuality and comparison assessment by forensic document examiners.

Key References in the Executive Summary

Boulanger-Carey M., Chen, H.T., Descloux, A., Ingle, J.F. and Park, K.I. (1984). "1982/83 End Office Connections Study: Analog Voice and Voiceband Data Transmission Performance Characterization of the Public Switched Network," *AT&T Bell Laboratories Technical Journal*, **63**(9): 2059-2119.

Boulanger, M., Johnson, M.E., and Vastrick, T. (2014). "Development of an Extensive Forensic Handwriting Database—Statistical Components," Topics in Statistical Simulation, Papers from the Seventh International Workshop on Statistical Simulation, Rimini, Italy, May 21-25, 2013.

Huber, R. and Headrick, A. (1999). *Handwriting Identification: Facts and Fundamentals*. CRC Press, Boca Raton, FL.

ISO/TR 14468:2010(E), Selected Illustrations of Attribute Agreement Analysis, Geneva: International Standards Organization.

NRC (2009). *Strengthening Forensic Science in the United States: A Path Forward*. National Research Council, Washington, DC, Committee on Identifying the Needs of the Forensic Sciences Community and Committee on Applied and Theoretical Statistics.

Srihari S.N., Cha, S.H., Arora, H., Lee, S. (2002). "Individuality of Handwriting", *Journal of Forensic Sciences*, **47**: 1-17.

Vastrick, T. (1998). "The Uniqueness of Handwriting", *Journal of American Society of Questioned Document Examiners*, **1**(1): 4-7.

Measuring the Frequency Occurrence of Handwriting and Hand-Printing Characteristics

1. Introduction

Forensic evidence has come under scrutiny in the past several years. Fienberg (2007), Finneran (2003), and Kennedy (2003) drew attention to some of the problems with the scientific underpinnings of forensic science in general. A pattern has emerged within the specialty discipline of forensic document examination involving admissibility rulings in which judges were admitting forensic document examination as reliable under the Daubert guidelines but chastising the discipline for having limited empirical bases. At a Questioned Documents Symposium in Ames, Iowa in 2008, investigators Vastrick and Schuetzner along with Forensic Document Examiners Kathleen Storer and Karen Runyon met and developed an outline of what was to become this research project.

Publication of the National Research Council report, *Strengthening Forensic Science in the United States: A Path Forward* (2009) further established that forensic science in general needed additional scientific underpinnings. Included in the report was a section on handwriting analysis for which these limitations were identified. In particular, this report noted with respect to handwriting analysis the following:

> "The scientific basis for handwriting comparisons needs to be strengthened. Recent studies (Kam et al, 1997) have increased our understanding of the individuality and consistency of handwriting and computer studies (Srihari et al, 2002) suggest that there may be a scientific basis for handwriting comparison, at least in the absence of intentional obfuscation or forgery. Although there has been only limited research to quantify the reliability and replicability of the practices used by trained document examiners, the committee agrees that **t**here may be some value in handwriting analysis."

Recognizing deficiencies in empirical bases, the National Institute of Justice launched several research initiatives to address these issues. Based on the groundwork performed prior to 2009, The National Center for Forensic Science at the University of Central Florida received a $550,000.00 grant to establish frequency occurrence proportions for specific cursive and hand printed characteristics by writers of diverse demographic backgrounds representative of the U.S. population. The proportions are designed for use by research investigators, forensic document examiners in their daily work, and forensic document examiners in courtroom situations during which questions of empirical bases is raised. The potential applications of the research and its results described in this report include use in both civil and criminal cases. The objectives of the project in the original proposal were, as follows:

1. Develop statistically valid frequency occurrence proportions of characteristics of handwriting and hand printing based on specimen samples from throughout the United States;
2. Provide practitioners of forensic document examination with statistical basis for reliability and measurement validity to accurately state their conclusions and assess complexity; and
3. Provide courts with the reliable data needed to understand the underlying scientific basis for the examinations and the conclusions.

A significant number of handwriting and hand printing characteristics were selected based on the belief that these characteristics were objective in nature – avoiding terms like "long stroke" or "sharp curve" that are susceptible to subjective interpretations from different classifiers. Approximately 2500 characteristics were initially selected but were eventually pared down to 786 characteristics. A multi-level sampling approach (Attribute Agreement Analysis) was utilized to gather handwriting and hand printing specimens. Over 1500 writers provided specimens to the study. The primary goal of this project is to determine how many of the writers did or did not utilize the selected handwriting and hand printing characteristics in their writing specimens. Using standard statistical sampling methodologies, frequency occurrence proportions for each characteristic within the general population can be determined within an established error rate for the sampling methodology. The utility of this database is demonstrated in a results section that addresses the validation of the concept of heterogeneity of handwriting and the application of the product rule in handwriting assessments among other statistical evaluations. The resulting data can be used to fulfill the requests of various courts for more scientific bases for probabilistic uniqueness and various uses in our daily examinations.

Throughout the project, strict adherence to standards has been paramount. This project is a statistical study, not a forensic study. As such, the methodologies used are driven by standards and best practices from within the statistical field. The resulting data from this project has the potential for significant effects on the profession of forensic document examination. One of the basic axioms of handwriting comparison is that no two writers utilize the exact same set of handwriting characteristics. The database created by this research and the resulting frequency occurrences provides the forensic and judicial communities some empirical data concerning this axiom.

Forensic document examination standards benefit from this data by strengthening their foundations such as providing additional basis for estimating level of confidence in stating conclusions for forensic document examinations as illustrated in Table 1. This table is drawn from ASTM-E1658 (2008) as specifically referenced in NRC Report (2009) and now codified through SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners (2012). Although the opinions in brief coupled with their elaborations provide an estimation of confidence, a quantification of these levels would be a welcome addition to the forensic and judicial communities. The proportions developed in this project potentially can serve to provide these categories with supporting calculations.

Examiner Conclusion	Elaboration
Identification	A definite conclusion that the questioned writing and known specimens are from a common source
Strong probability (highly probable, very probable)	Evidence is persuasive, yet some critical quality is missing. "Virtually certain".
Probable	Points strongly towards identification but falls short of the "virtually certain" degree of confidence
Indications (evidence to suggest)	There are a few significant features for handwriting comparison purposes, but those features are also in agreement with another body of writing
No conclusion (totally inconclusive, indeterminable)	There are limiting factors (e.g., disguise) or lack of comparable writing so that the examiner does not favor one opinion over another; zero point on the confidence scale
Indications did not	There are a few features which are of significance for handwriting comparison purposes, but those features are in agreement with another body of writing. This opinion comparable in strength to Indications (evidence to suggest)
Probably did not	Evidence is quite strong against a common writer but falls short of "virtually certain"
Strong probability did not	Virtual certainty against a common writer
Elimination	Highest degree of confidence—no doubt that the questioned and known documents written by different individuals

Table 1 - Current Gradations of Forensic Document Examiner Opinions (ASTM-E1658, (2008); SWGDOC (2012))

From basic textbooks to recent research in forensic document examination, authors have applied or suggested the application of the product rule in estimating overall heterogeneity of handwriting. With the data collected in this project, the raw data of frequency occurrence proportions of multiple characteristics will be compared to the product rule result applied to those same characteristics in an effort to determine whether a pattern of the parallel results being statistically different (acknowledging the hypothesis that handwriting characteristics are interdependent) or a pattern of the parallel results not being statistically different

(rejecting the hypothesis that handwriting characteristics are interdependent) exists. Moreover, qualitative opinions as to the nature of handwriting features as a function of age, gender, ethnicity, education and handedness will be explored for a variety of features.

2. Literature Review

The scientific literature describing various sources that affect handwriting is conveniently summarized in the compendium by Huber and Headrick (1999). This reference captures the state of the art as of 1999 and suggested demographic along with other factors to be considered. Huber and Headrick note both extrinsic and intrinsic factors that influence handwriting. The following are specific extrinsic factors that they noted:

Writing systems (national, cultural and occupations)
Physiological constraints
- *Foot and mouth holding the writing instrument*
- *Artificial aids (prostheses)*
- *Sightedness*

Gender
Normal Physical Attributes
- *Maturity, practice, development*
- *Handedness (left/right)*

Physical abnormalities in health
- *Handwriting as a diagnostic tool*
- *Organically related illnesses*
- *Medications*

Infirmity
- *Guided hands*
- *Senility*

Mental state
- *Emotional stress, nervousness, instability*

Injury (unrelated to hands)

For intrinsic factors, the following are noted:

Imitation (disguised or forged)
Circumstantial
Temporal states (non-chronic)
- *Alcohol*
- *Hallucinogens or hard drugs*
- *Fatigue and physical stress*

Literacy and Education

There has been significant research into computer analog and statistical evidence in support of the heterogeneity of handwriting by Srihari and colleagues (2002, 2003, 2008, 2010, 2013). Likewise, there is ample research concluding that

forensic document examiners perform one of their principle duties of handwriting comparison more accurately than equally educated but untrained subjects to a statistically significant level (Kam and colleagues—1994, 1997, 1998, 2001, 2010; Sita, Found and Rogers, 2002; Dyer, Found and Rogers, 2006).

However, one would be remiss to consider the above as the totality of research performed on the subject of the heterogeneity of handwriting and hand printing. Classification systems for handwriting and hand printing have been available for almost 100 years and in not one instance has any of these classification systems found indistinguishable handwriting from two different sources (Lee and Abbey (1922); Popkiss and Moore (1945); U.S. Secret Service and Bundeskriminalamt (FISH); Huber (2000); Livingston (1959) and Nicholson (1999)). Additional publications addressing probabilities and uniqueness include Osborn (1929); Muehlberger, et al (1977); and Hilton (1958).

This study is not the first on the subject of frequency occurrence in handwriting. Other treatises include Bishop (2012); Chamberland and Ghirotto (1990); Eldridge et al (1984); Horton (1996); Huber (2000); Kelly (2002); Livingston (1963); Savoie (2011); Shiver (1996); Vastrick (1998) and Welch (1996). Research has also studied the handwriting of close sub-groupings such as twins (Beacom (1960); Gamble (1980); Boot (1998)) (Srihari (2008)); adolescents (Cusak and Hargett (1988); and schoolmates (Durina (2009). Vastrick (1998) lists 27 different sub-grouping examinations or studies conducted by forensic document examiner respondents with a total of 1,490 man-years of experience to a questionnaire on the subject.

3. Methods

The basic premise of this project is simple. Take handwriting specimens from writers selected to constitute a "representative" sample. Examine each set of writing and note the presence or absence of predetermined handwriting characteristics. Total up the numbers and divide by the number of participants to obtain a frequency occurrence proportion. However, the complexity of the selection process of writers and the magnitude of the characterization effort of their writing, along with the pilot testing necessary for each step of the methodology provided a challenging effort to the research team, particularly keeping in mind the required use of standard methodologies and best practices.

Absence of a Sampling Frame and Acquisition of Samples. In many sampling applications, a sampling frame is available which consists of all units in the relevant population. In this project a sampling frame was not available. No frame of all possible individuals from which a random sample could be drawn was discovered. The target population for this study was defied as adults 18 years old or older who are capable of providing writing samples by hand and in English (preferably both cursive and hand printed forms). Younger writers are increasingly less trained in cursive writing, so some specimens provided are strictly the hand printed version.

Excluded are adults who are not capable of handwriting (i.e., foot or mouth writing), blind, senile or require a guided hand to write.

The rationale for our approach was described by Boulanger, Johnson and Vastrick (2013), as follows:

> "Thus our approach to data collection changed from a probabilistic sampling process to the development of a data collection process that will lead to a large sample of "writers" deemed representative of the target population. The approach we followed was based on a study done to evaluate the performance of the national telecommunications network before the breakout of the monopoly service provider, AT&T (Boulanger Carey et al, 1999; ASTM-E105-04, 2004; ASTME141-91, 1991). There, as in our situation with handwriting, it was not possible to construct a sampling frame of all the potential telecommunication paths in the US and a multi-level sampling approach based on identification of strata and clusters was developed to lead to a quasi-representative sample."

The planning approach used had seven steps governing the data collection process:

1 Research factors influencing handwriting
2 Define stratification variables based on key factors influencing handwriting and key variables describing the target population
3 Define strata for selected variables
4 Estimate proportions within the strata of the target population
5 Define the data collection process for obtaining a representative sample of the target population, deemed so by meeting the quota guidelines
6 Provide guidelines for the data collection process to the data collectors
7 Audit the data collection process for adherence to the plan and for quality control

The specific handling of the factors effecting handwriting as noted by Huber and Headrick in terms of inclusion in our target is summarized in Table 2.

In the absence of a sampling frame, the investigators opted to obtain samples from a collection of adults who are reasonably representative of the target population regarding demographics and other factors known to influence handwriting as indicated in Table 2. In order to obtain the set of specimens corresponding to a representative sample of writers, a large set of specimens were obtained and then at the analysis stage, a subset of the population enjoying common overall demographic characteristics to the total population was determined for the ultimate determination of individual and multi-dimensional frequencies. The paring process was conducted by the statisticians using appropriate methodologies.

FACTORS INFLUENCING HANDWRITING (Huber & Headrick)				OUR SAMPLING PROCESS	
Section in H&H	**Reference in H&H**	**Extrinsic or Intrinsic factor**	**Variables in Huber and Headrick**	**How handled in our sample?**	**Information requested**
Extrinsic Factors					
8.37	A	E	Writing systems: National, cultural, and occupational	People who are in the US (including foreigners traveling), able to write in English (not necessarily speaking English)	Location of 3rd grade schooling
8.37	B	E	Physiological constraints:		
	B1		Foot & Mouth	Do not accept in sample	
	B2		Use of artificial aids (prostheses)	Accept in sample	Do you have any physical imparities or injuries?
	B3		Deafness and/or sightlessness	Do not include blind people in sample, accept deaf people if communication is possible	Do you have any physical imparities or injuries?
8.37	C	E	Genetic factors: Sex	Familial relationship and multiple births are of no interest in this study. Ignore	Record Gender
8.37	D	E	Physical (Normal)		
	D1		Maturity, Practice, and Development	Accept only people 18 years old or older	Record age
	D2		Handedness	No control of which hand should write in the case of ambidexterity.	Record hand doing the writing. Do not record Grasp
8.37	E	E	Physical (Abnormal state of health)		
	E1		Handwriting as a diagnostic tool	Not relevant to our study - ignore	
	E2		Illnesses organically related	Accept in sample	Do you have any physical imparities or injuries?
8.37	F	E	Medications	No control, no asking	No information asked
8.37	G	E	Infirmity		
	G1		Senility	Do not accept in sample	
	G2		Guided hands	Do not accept in sample	
8.37	H	E	Mental State of Writer (Emotional stress, nervousness, instability)	Accept in sample	No information requested or noted
8.37	I	E	Injury	Accept in sample	Do you have any physical imparities or injuries?
Intrinsic Factors					
9.38	A	I	Imitation	Not relevant to our study - ignore	
9.38	B	I	Circumstantial	Control environment	Provide pen and paper; Provide "comfortable position for the person to write with adequate support level"
9.38	C	I	Temporal states of the writer		
	C1		Alcohol	Accept in sample	
	C2		Hallucinogens and hard drugs	Accept in sample	
	C3		Hypnosis	Accept in sample	
	C4		Fatigue & physical stress	Accept in sample	Flip-flop printing and cursive writing - record order
9.38	D	I	Literacy and Education		Record information on education level

Table 2 - Handling of Factors Effecting Handwriting

To achieve representativeness, stratification variables were established in line with the most relevant demographic and handwriting factors as suggested by Huber and Headrick. Table 3 provides the ranges initially set for the collection of specimens.

Forensic document examiners and students collected samples in accordance with these protocols. The overall characteristics of the specimen providers were roughly in proportion with the characteristics in the target population (with the proportions oscillating in the course of the collection).

Reference In Huber and Headrick	Stratification variable	Strata definition	Strata proportion in US	Minimal Goal Specification (80% per factor)
A	Writing Systems	Location 3rd schooling In US	80%	> 70%
		Location 3rd schooling NOT In US	20%	>10%
C	Gender	Male	49.0%	>40%
		Female	51.0%	>40%
D	Age	18-30	33.0%	>20%
		> 30 up to 50	36.0%	>30%
		> 50	41.0%	>30%
D	Handedness	Right	90.0%	>75%
		Left	10.0%	>5%
C	Temporal State	Night (after 8pm)		>20%
		Day (before 8pm)		>60%
D	Education	High School or less	49.0%	>30%
		Beyond High School	51.0%	>50%
N/A	Race	White	63.7%	>55%
		Black	12.6%	>10%
		Hispanic	16.3%	>11%
		Asian	4.8%	>4%
N/A	US Region (where samples were taken)	North West		>15%
		North East		>15%
		Midwest		>15%
		South West		>15%
		South East		>15%
N/A	Location	College and universities		>20%
		Religious places		>20%
		Social and non-social gathering areas (e.g., malls, night entertainment, jury waiting rooms, restaurants, fast food places)		>40%

Table 3 - Final Data Collection Goals

The specimen collectors were provided general guidance as to the categories of participants in the specimen collections. The intent was to meet a fixed minimal goal of 80% of the sample for each stratification variable, as given in Table 4. Precise quotas were recognized as likely unachievable, so the last 20% without any quota offered a cushion and some flexibility in collecting samples. This also

implicitly introduced some degree of randomness into the collection process to possibly account for unforeseen factors. In the absence of a probabilistic sampling approach, such inadvertent randomness is welcome if not a panacea. We also suggested that the collectors choose their collection locations according to the following preliminary plan to induce some representativeness:

20%: Universities (young adults; education at least high school with some beyond, foreign adults)
20%: Worship locations (mature adults) at churches, temples, etc. (various races and ethnicities dominant)
20%: Evening entertainment locations after 8pm
20%: Restaurants and fast food locations (education less than high school)
20%: Surveyor/data collector discretion

The collectors selected the location with an appreciation for the ranges in our study. Any information that the collector could provide for potential adjustment at the analysis stage was encouraged and welcome. In summary, great lengths were taken to ensure that the samples collected would satisfy the representativeness for our project sample with respect to the target population while at the same time minimizing any negative effects of pure quota or pure random data collection.

The specimen form utilized was a slightly modified form utilized by Dr. Sargur Srihari in previous research with his gracious approval. The form embodies each character in different placements within a word yet of a manageable length. The form is illustrated in Appendix 1.

In an effort to prevent unintentional variations in the entire collection process all handwriting specimen forms were printed at one time from one source. In addition, a bulk supply of non-retractable ballpoint pens were purchased and distributed for use in specimen collection. Project management worked with the UCF Institutional Review Board to assure compliance with federal regulations involving human subject participation. In particular, contributors who decided to withdraw their participation and specimens part way through the effort were free to do so. The participants providing specimens were volunteers who responded to the societal benefit by contributing to a research project at the university. No payments were made to any participants providing specimens.

Pilot Studies. An initial step of this project was to identify and define multiple characteristics of letters (cursive and hand printed), numbers and punctuation marks then prepare a database for classifiers to determine the presence or absence of these features. A significant number of handwriting and hand printing features were selected based on the belief that these characteristics were objective in nature – avoiding terms like “long stroke” or “sharp curve” that are susceptible to subjective interpretation from different classifiers. These initial characteristics were selected by a team of forensic document examiners who were assigned a series of about five letters and a few numerals and punctuation marks each. Some common sources of literature for initial selection included Schuetzner (1999, 2000)

and common systems of handwriting utilized in schools within the United States. The examiners were instructed to create a list of features that were, based on their education, training and experience,, objective in nature. In addition, each examiner was to provide, as best as possible, a comprehensive list of features. Investigator Vastrick created and scanned illustrative images to support the feature descriptions. Vastrick and Investigator Schuetzner reviewed the feature descriptions and reached consensus regarding consistent terminology as prescribed by standards and common texts' description of applicable nomenclature. The initial list and images were submitted to the database specialist and a database was created that used a checkbox system for feature presence/absence. Over 2500 individual characteristics were defined initially.

An Attribute Agreement Analysis (AAA) is an accepted statistical method for evaluating the level of agreement in answers among multiple participants or the same participant at different times. An AAA was designed and implemented to test each characteristic for objectivity and reproducibility of results. The AAA also simultaneously tested the method of presentation (the database) for the same objectivity and reproducibility of results. The statistical aspects of the attribute agreement analysis approach used is embodied in ISO TR 14468 and supported by the international statistical community. The design of the attribute agreement analysis was to have three ABFDE-certified forensic document examiners classify 5 complete specimens (both cursive and printed) including two replicates for each of the original 2500 characteristics. The specimens ranged from highly-skilled, extremely neat to lower-skilled, nominally clear and can be considered representative of the range of writer skill levels present within the specimens available at the time of the study. Designating the three examiners as A, B and C, the examiners considered the specimens, as follows:

Cursive Letters (upper and lower case)
- Examiner A: #4, #7, #11, #201, #222 and replicated #4, #7
- Examiner B: #4, #111, #201, #222 and replicated #4, #111, #222
- Examiner C: #4, #7, #95, #111, #201, #222 and replicated #7

Printed Letters (upper and lower case)
- Examiner A: #4, #7, #11, #201, #222 and replicated #4, #222
- Examiner B: #4, #111, #201, #222 and replicated #4, #111, #222
- Examiner C: #4, #7, #95, #111, #201, #222 and replicated #7

A total of 903 characteristics survived the attribute agreement analysis described in this section. Subsequent paring resulted in a final project total of 786 characteristics. These features are listed as part of the proportion results table in Section 4 of this report. The surviving characteristics had one-hundred percent agreement by multiple examiners across multiple handwriting specimens, including replicated reviews by the same examiner. A single disagreement was grounds for dismissal as this instance would shed doubt on the reliability of the data. There was no analysis as to the reasons for rejection of the approximately 1600 features.

Presence/absence of a feature is the sole "measurement" per characteristic, unlike attribute agreement studies involving physical laboratories in which continuous measurements are likely to be collected in such studies. Given the large number of features, a simple presence/absence response was in order for the hundreds of features inspected. A benefit of the attribute agreement analysis was the elimination of characteristics that generated discrepant opinions and to produce a list of features appropriate for assessment over the main 1500 specimens collected.

An AAA was undertaken by Investigator Vastrick at the 69th Annual Meeting of the American Society of Questioned Document Examiners in order to test the feasibility of using handwriting management software. The test encompassed the use of commercially-available handwriting management software to separate out each occurrence of each character. As such, the classifier would have a page of paper with all lower case "a"'s on it that were present in the handwriting specimens; they would also get a page with all lower case "b"'s on it. The perceived advantage would be to have all versions of a particular character conveniently placed side-by-side for review. The perceived disadvantage was that the classifier was using a reproduction of modest resolution. Approximately 50 forensic document examiners spent one-half day classifying numerous handwriting specimens which had been processed by the handwriting management software. Based on the results of the AAA pilot study, Investigators Boulanger, Johnson and Vastrick collectively decided that neither handwriting management software nor the use of photocopies was sufficiently reliable for this project. The use of original handwriting is a Best Practice within forensic document examination. In place of handwriting management systems an assisting illustration independent of the specimens was devised that provided classifiers with the location of each specific character throughout the text of the handwriting form to help in locating every example of any given character. Appendix 2 contains examples of the assisting illustration

All handwriting specimens were collected and submitted to the National Center for Forensic Science (NCFS). Specimens were collected from each region of the country and numerous locations within each region. More importantly the specimens were from participants who received their early elementary education in 47 states. Each specimen was provided a writer number which was subsequently used for any computer referencing. The biographical data and other extrinsic/intrinsic collected factors were recorded and each form was placed in a document protection sheet. The NCFS then distributed the specimens to the various classifiers for classification and data entry into the database. Upon completion of the analysis the forms were returned to the NCFS where they are currently being stored per government privacy requirements.

The database was developed using a common commercially-available database software program modified specifically for this project. Forensic document examiners provided a combination of characteristic descriptions with accompanying example images that the database expert, Heather Burske incorporated into the user-friendly database (Figure 1).

The specimens were classified by forensic document examiners and upper division forensic science majors at the University of Central Florida (UCF). Classification does not require a comparison process or an associative or dissociative conclusion which is the proper venue of a qualified forensic document examiner. Classifiers were required to determine the presence or absence of defined basic features that are both described and illustrated to which end these students could be adequately trained. Investigator Vastrick provided training for all classifiers participating in this project. Student participants were required to have a recommendation from the head of the Forensic Science curriculum at UCF and successfully complete an interview with Investigator Vastrick. Some students chose to receive course credit in research methodology from UCF for their efforts and some students were paid a nominal fee. Written material to include the assisting illustration form was also provided. Quality control was maintained by Investigator Vastrick's periodic and random review of the student work product.

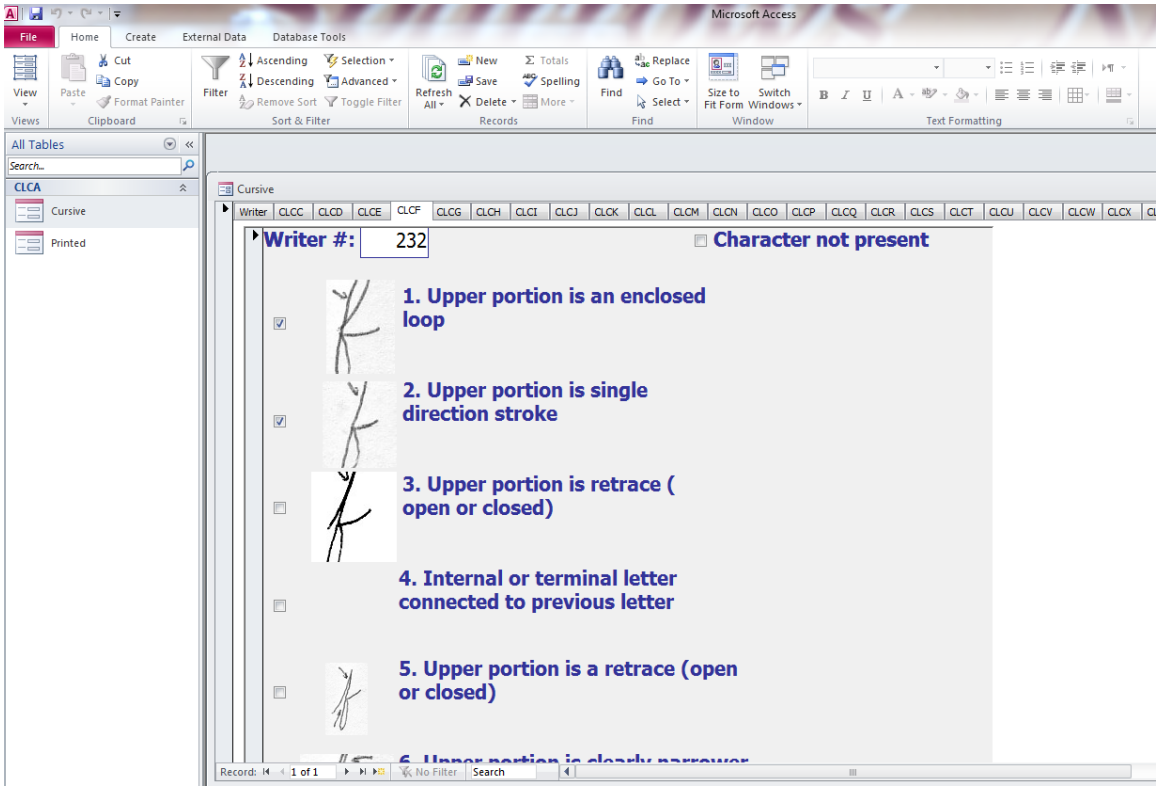

Figure 1 – Screenshot of project database illustrating checkbox format, feature descriptions and accompanying illustrations. The complete database is available online. Contact Thomas Vastrick at vastrick@yahoo.com for locations.

There were no blind studies affiliated with this project. Blind studies are used for cause-and-effect processes. This project has no cause-and-effect studies. As such, it would be contrary to best practices to incorporate a process that is not designed for the form of study being conducted.

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

4. Results

Before producing the main numerical results from this project, considerable effort was expended in cleaning the data—so-called data preparation. Appendices 3 and 4 describe in detail the steps taken to get the original data set cleansed of difficulties. The original spreadsheet of 1517 specimens was pared down by eliminating specimens of which demographics had been provided but the specimens had not yet been examined. For subsequent analyses, it was also critical to code the demographic variables in a unique and consistent fashion. For example, over three hundred different responses were recorded as to the location of the writers' 2nd/3rd grade education. These set of results were reduced to US, non-US and unknown according to the recoding scheme outlined in Appendix 3.

In examining the demographics associated with the 1517 specimens, we concluded the following:

1. Slightly under-represented in the 30-50 age category.
2. Under-represented with males.
3. Considerably over-represented with Caucasians.
4. No problem with handedness.
5. Considerably under-represented with high school or less education.
6. Considerably under-represented with contributors who received their 2nd/3rd grade educations outside the US.

Thus, using the entire set of specimens would yield a population of contributors which collectively do not represent the population of interest with respect to demographics. However, we can carefully pare down to a *subset* of the specimens in order to attain closer correspondence between the actual and stratified percentages while balancing this objective with the goal of having as large a sample as possible.

The following decisions were made to create as large a sample of specimens as possible, from the set of 1517 specimens while improving if not exceeding the quota percentages:

1. Include all specimens corresponding to at most high school level education.
2. Include all specimens corresponding to those contributors having their 2nd/3rd grade education outside the US.
3. Include all non-Caucasians.
4. Include all males.
5. Eliminate 17 duplicate specimen reviews discovered at this stage in the process.
6. Eliminate the 56 specimens that are in the full file but for which there are no features examined.
7. Eliminate an additional 19 specimens for which a TRUE is given for "not present" for each character and then FALSE is the response for each

feature, leading to 19 duplicate specimens in terms of the responses. These specimens are de facto not examined.

With these rules in place, a total of 880 cursive specimens and 839 hand printed specimens resulted. Although achievement of all of the goal bounds originally set, we have improved the representativeness of the population of contributors greatly. Checking the major categories for this subset of the population yields the following results for the cursive specimens:

Age	count	%	Quota %
24	332	38.2%	> 20%
40	243	27.9%	> 30%
60	295	33.9%	> 30%
Unknown	10		

Table 4 – Cursive Population Sampling Based on Age

Gender	count	%	Quota %
Female	390	44.6%	> 40%
Male	484	55.4%	> 40%
Unknown	6		

Table 5 – Cursive Population Sampling Based on Gender

Ethnicity	count	%	Quota %
African-American	95	11.4%	> 10%
Asian	49	5.9%	> 4%
Caucasian	612	73.4%	> 55%
Hispanic	83	10.0%	>11%
Mixed	3	0.4%	
Native American	3	0.4%	
Other	5	0.6%	
South Pacific	7	0.8%	
Unknown	23		

Table 6 – Cursive Population Sampling Based on Ethnicity

Right or Left	count	%	Quota %
Left	78	9.0%	> 5%
Right	791	91.0%	> 75%
Ambidextrous	3		
Unknown	8		

Table 7 – Cursive Population Sampling Based on Handedness

Education Level	count	%	Quota %
High School or less	279	32.4%	> 30%
HS plus	582	67.6%	> 50%
Unknown	19		

Table 8 – Cursive Population Sampling Based on Level of Education

2nd/3rd Grade Education Location	Count	%	Quota %
US	626	89.9%	> 70%
Not US	70	10.1%	> 10%
Unknown	184		

Table 9 – Cursive Population Sampling Based on Location of Early Education. Forty-seven of the fifty states were represented in the sample collection.

Restriction to the 880 cursive specimens yields demographics that meet the prescribed quotas with just two exceptions—the age category for 30-50 years is 27.9% in the proposed final sample that is slightly below the prescribed 30% and the Hispanic category is 10.0% versus the prescribed 11%. Any further additions to the membership of the final sample from the existing set of specimens will adversely affect one of the other quotas. The restricted set of 880 cursive specimens provides the basis for the frequencies of characteristics reported in Table 16. The subsequent results in this report are based on analyses of the 880 specimens, collectively designated the cursive project sample.

Age	count	%	Quota %
24	321	38.8%	> 20%
40	231	27.9%	> 30%
60	276	33.3%	> 30%
Unknown	11		

Table 10 – Hand Printed Population Sampling Based on Age

Gender	count	%	Quota %
Female	336	40.3%	> 40%
Male	497	59.7%	> 40%
Unknown	6		

Table 11 – Hand Printed Population Sampling Based on Gender

Ethnicity	count	%	Quota %
African-American	94	11.5%	> 10%
Asian	48	5.9%	> 4%
Caucasian	572	70.1%	> 55%
Hispanic	87	10.7%	>11%
Mixed	3	0.4%	
Native American	3	0.4%	
Other	4	0.5%	
South Pacific	5	0.6%	
Unknown	23		

Table 12 – Hand Printed Population Sampling Based on Ethnicity

Right or Left	count	%	Quota %
Left	71	8.6%	> 5%
Right	757	91.4%	> 75%
Ambidextrous	4		
Unknown	7		

Table 13 – Hand Printed Population Sampling Based on Handedness

Education Level	count	%	Quota %
High School or less	282	34.3%	> 30%
HS plus	541	65.7%	> 50%
Unknown	16		

Table 14 – Hand Printed Population Sampling Based on Level of Education

$2^{nd}/3^{rd}$ Grade Education Location	Count	%	Quota %
US	637	90.4%	> 70%
Not US	68	9.6%	> 10%
Unknown	134		

Table 15 – Hand Printed Population Sampling Based on Location of Early Education. Forty-seven of the fifty states were represented in the sample collection.

Restriction to the 839 hand printed specimens yields demographics that meet the prescribed goals with two exceptions—the age category for 30-50 years is 27.9% in the proposed final sample that is slightly below the prescribed 30% and the Not US category is 9.6% versus the prescribed 10.0% for the education location. Any further additions to the membership of the final sample from the existing set of specimens would adversely affect one of the other quotas. The restricted set of 839 specimens provides the basis for the frequencies of characteristics reported in Table

5. The subsequent results in this report are based on analyses of the 839 specimens, collectively designated the Study Printed Sample.

A key deliverable in this project is an estimation of the frequencies of the presence/absence of features within the cursive project sample. There were 435 specific cursive characteristics examined for the cursive specimens, covering 49 lower case and upper case letters. The lower case letters "a", "b" and "d" did not have any features that survived the attribute agreement analysis described previously. There were 351 specific hand printed characteristics examined for the hand printed specimens, covering 50 lower case and upper case letters. The lower case "c" and "i" did not have any features that serviced the attribute agreement analysis. The full set of proportions of presence of the features is given in Table 4.

The proportions can be very close to zero (rare) or one (common). Figures 2 and 3 illustrate histograms of the observed proportions of presence in the project samples. Some specific instances of characteristics will be further examined in the context of some research questions.

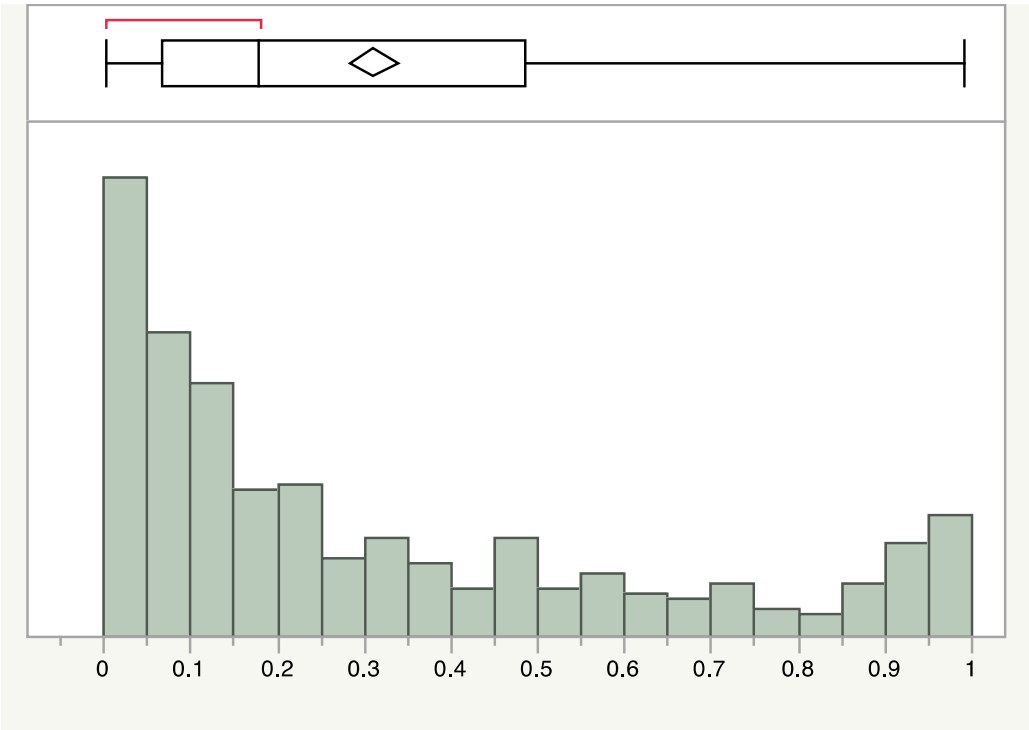

Figure 2. Histogram of Features Present in the Cursive Project Sample.

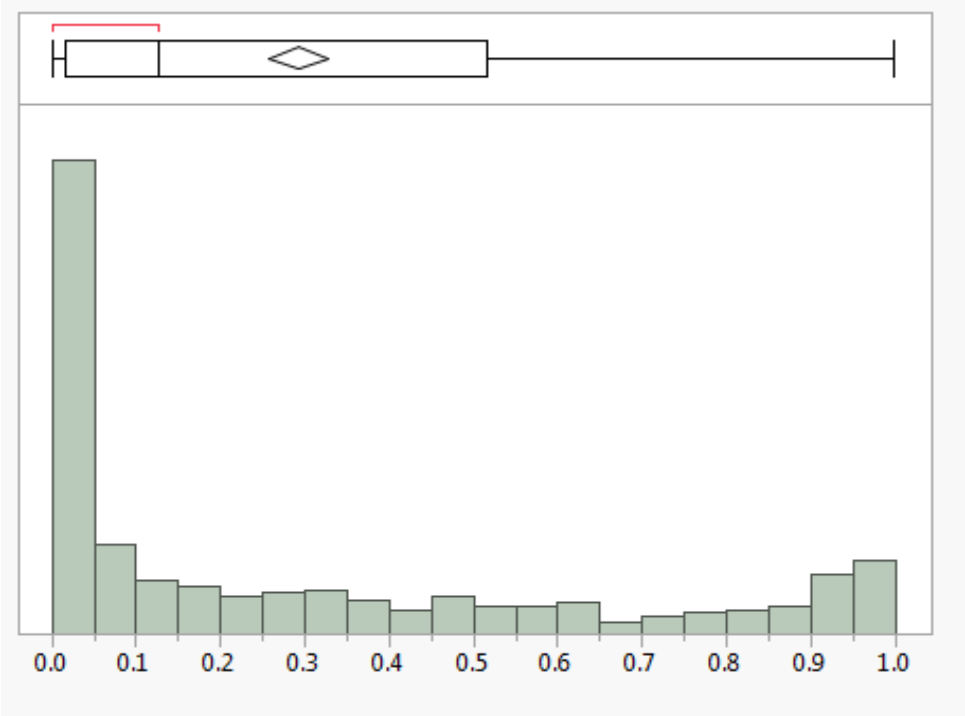

Figure 3. Histogram of Features Present in the Hand Printed Project Sample

Table 16 contains frequency occurrence proportions for each of the selected cursive characteristics. The feature terms are abbreviated. "CLC" represents "Cursive Lower Case" thus "CLCF" represents "Cursive Lower Case "f" and "CLCR" represents "Cursive Lower Case "r". Similarly "CUC" represents "Cursive Upper Case. Therefore, "CUCA" represents "Cursive Upper Case "A".

FEATURE	DESCRIPTION	COUNT	FREQUENCY
CLCC 1	Character not present		
CLCC 2	1. Internal or terminal letter connected to previous lower case letter	865	0.94451
CLCC 3	2. Cap is enclosed retrace	865	0.88208
CLCC 4	3. Internal or terminal letter not connected to previous letter	865	0.22197
CLCC 5	4. Width to height ratio imbalance taller than wide or wider than tall is not obvious	865	0.86012
CLCC 6	5. Internal c not connected to both previous and subsequent letters	865	0.10058
CLCE 1	Character not present		
CLCE 2	1. Letter is enclosed loop	877	0.99088
CLCE 3	2. Internal or terminal letter connected to previous lower case letter	877	0.97834
CLCE 4	3. Internal e not connected to both previous and subsequent letter	877	0.14937
CLCE 5	4. Initial or internal letter connected to subsequent letter	877	0.97149
CLCF 1	Character not present		
CLCF 2	1. Upper portion is an enclosed loop	862	0.79814
CLCF 3	2. Upper portion is single direction stroke	862	0.15661
CLCF 4	3. Upper portion is retrace (open or closed)	862	0.45708
CLCF 5	4. Internal or terminal letter connected to previous letter	862	0.89791
CLCF 6	5. Upper portion is clearly narrower than lower portion	862	0.50928
CLCF 7	6. Slant of the upper and lower portions is approximately the same	862	0.67285
CLCF 8	7. Lower portion is a down stroke only	862	0.08933

CLCF 9	8. Lower portion is curved or loop clockwise	862	0.14153
CLCF 10	9. Terminal portion touches staff	862	0.86079
CLCG 1	Character not present		
CLCG 2	1. Descender is curved counterclockwise	866	0.0739
CLCG 3	2. Descender is enclosed loop, triangulation or other similar design	866	0.95266
CLCG 4	a. Lower loop/design intersects upper loop	866	0.23672
CLCG 5	3. Internal g not connected to both previous and subsequent letter	866	0.13395
CLCG 6	a. Terminus is clearly sloped downward	866	0.01386
CLCG 7	b. Terminus is sloped approximately horizontal	866	0.05427
CLCG 8	4. Initial or internal letter connected to subsequent letter	866	0.94111
CLCH 1	Character not present		
CLCH 2	1. Internal or terminal letter is connected to previous stroke	869	0.95627
CLCH 3	2. Initial stroke clearly begins to right of staff	869	0.04028
CLCH 4	3. Initial stroke begins approximately on staff	869	0.22209
CLCH 5	4. Staff is enclosed loop	869	0.27503
CLCH 6	5. Internal h not connected to both previous and subsequent letter	869	0.13119
CLCH 7	6. Initial or internal letter connected to subsequent letter	869	0.96318
CLCH 8	7. Initial/ internal letter not connected to subsequent letter	869	0.21174
CLCI 1	Character not present		
CLCI 2	1. Internal or terminal letter connected to previous lower case letter	878	0.96697
CLCI 3	2. Staff is retrace (open or closed)	878	0.9738
CLCI 4	3. I-dot is present	878	0.9738
CLCI 5	a. Dash straight (horizontal, vertical, or diagonal)	878	0.50797
CLCI 6	b. Circle	878	0.04214
CLCI 7	c. Other	878	0.57062
CLCI 8	d. I-dot is connected to body	878	0.01595
CLCJ 1	Character not present		
CLCJ 2	1. Internal or terminal letter connected to previous lower case letter	865	0.91329
CLCJ 3	2. Initial stroke begins at top of staff	865	0.2578
CLCJ 4	3. Peak is enclosed loop	865	0.10405
CLCJ 5	4. Peak is rounded	865	0.0763
CLCJ 6	5. Dot is not present	865	0.26127
CLCJ 7	6 .Dot is present	865	0.90636
CLCJ 8	a. Dot or approximate dot (not clearly a dash)	865	0.72023
CLCJ 9	b. Clearly dash – straight	865	0.29711
CLCJ 10	c. Other	865	0.11908
CLCJ 11	d. I-dot is not clearly aligned to either side of staff	865	0.46474
CLCJ 12	7. For initial and internal letter descender is enclosed loop with ending stroke	865	0.09133
CLCJ 13	8. For initial and internal letter descender is enclosed loop with connecting stroke to subsequent letter	865	0.82659
CLCJ 14	9. For initial and internal letter descender is approximately straight line	865	0.02775
CLCJ 15	10. For initial and internal letter descender is clearly curved (not enclosed loop) with ending stroke	865	0.09827
CLCJ 16	11. For initial and internal letter descender is other design with connecting stroke to subsequent letter	865	0.04393
CLCK 1	Character not present		
CLCK 2	1. Internal or terminal letter connected to previous lower case letter	871	0.94259
CLCK 3	2. Initial stroke begins at top of staff	871	0.29851
CLCK 4	3. Staff is enclosed loop	871	0.6946
CLCK 5	4. Internal k is not connected to both previous and subsequent letter	871	0.07922
CLCK 6	5. Initial or internal letter not connected to subsequent letter	871	0.17796
CLCL 1	Character not present		
CLCL 2	1. Initial stroke begins with clearly counterclockwise curve	875	0.05829
CLCL 3	2. Initial stroke begins with clearly clockwise cure	875	0.02171
CLCL 4	3. Initial stroke is other shape	875	0.48229
CLCL 5	4. Body is open loop	875	0.95886

CLCL 6	5. Body is retrace (open or closed)	875	0.39771
CLCL 7	6. Internal l is not connected to both previous and subsequent letter	875	0.11429
CLCM 1	Character not present		
CLCM 2	1. Internal or terminal letter connected to previous lower case letter	873	0.94502
CLCM 3	2. Left peak is clearly pointed	873	0.77892
CLCM 4	3. Left peak is enclosed loop	873	0.01604
CLCM 5	4. Left leg is enclosed loop	873	0.05727
CLCM 6	5. Middle leg is enclosed loop	873	0.02405
CLCM 7	6. Internal m not connected to both previous and subsequent letter	873	0.08362
CLCM 8	7. Initial and internal letter connected to subsequent letter	873	0.93242
CLCM 9	8. Initial or internal letter not connected to subsequent letter	873	0.15693
CLCN 1	Character not present		
CLCN 2	1. Internal or terminal letter connected to previous lower case letter	871	0.96441
CLCN 3	2. Left peak is clearly pointed	871	0.87945
CLCN 4	3. Left peak is open loop	871	0.02067
CLCN 5	4. Left leg is enclosed loop	871	0.06085
CLCN 6	5. Overcurve has clearly defined point	871	0.71412
CLCN 7	6. Internal n not connected to both previous and subsequent letter	871	0.12285
CLCN 8	7. Initial and internal letter connected to subsequent letter	871	0.95752
CLCO 1	Character not present		
CLCO 2	1. Internal or terminal letter connected to previous lower case letter	872	0.97018
CLCO 3	2. Internal or terminal letter not connected to previous	872	0.44954
CLCO 4	a. Initial stroke clearly begins left of center of letter	872	0.25459
CLCO 5	b. Initial stroke clearly begins right of the loop	872	0.03096
CLCO 6	c. Initial stroke is enclosed loop	872	0.01376
CLCO 7	d. Initial stroke clearly is curve	872	0.12959
CLCO 8	e. Initial stroke is approximately straight line	872	0.24197
CLCO 9	f. Initial stroke is extraneous/other	872	0.05619
CLCO 10	3. Loop is closed	872	0.97248
CLCP 1	Character not present		
CLCP 2	1. Staff is retrace (open or closed)	864	0.85764
CLCP 3	2. Width to height ratio imbalance is not obvious	864	0.83102
CLCP 4	3. End of loop touches staff	864	0.9375
CLCP 5	4. Internal p not connected to both previous and subsequent letter	864	0.10069
CLCQ 1	Character not present		
CLCQ 2	1. Internal or terminal letter connected to previous lower case letter	856	0.90187
CLCQ 3	2. Internal or terminal letter not connected to previous lower case letter	856	0.16706
CLCQ 4	3. Letter is standard design	856	0.91939
CLCQ 5	a. Cap is clearly open retrace with initial stroke on top	856	0.11332
CLCQ 6	b. Cap is clearly open retrace with initial stroke on bottom	856	0.13902
CLCQ 7	c. Top of staff is retrace (open or closed)	856	0.85748
CLCQ 8	d. Top of staff is other	856	0.08645
CLCQ 9	e. Lower extender is approximately straight single direction stroke	856	0.08879
CLCQ 10	4. Internal q not connected to both previous and subsequent letter	856	0.06893
CLCQ 11	5. Initial or internal letter connected to subsequent letter	856	0.87033
CLCQ 12	6. Initial or internal letter not connected to subsequent letter	856	0.18107
CLCR 1	Character not present		
CLCR 2	1. Double peak	863	0.76014
CLCR 3	2. Clearly pointed single peak	863	0.55968
CLCR 4	3. Internal r not connected to both previous and subsequent letter	863	0.14716
CLCS 1	Character not present		
CLCS 2	1. Design is printed/modified printed	868	0.20276
CLCS 3	2. Design is standard	868	0.95046
CLCS 4	a. Peak contains retrace (open or closed)	868	0.61521
CLCS 5	3. Internal s not connected to both previous and subsequent letter	868	0.09908

CLCT 1	Character not present		
CLCT 2	1. Internal or terminal letter connected to previous lower case letter	874	0.97025
CLCT 3	2. Staff is retrace (open or closed)	874	0.94851
CLCT 4	3. Cross bar is absent	874	0.12929
CLCT 5	4. Cross bar is to the right of the staff	874	0.15789
CLCT 6	5. Cross bar approximately bisects staff	874	0.9119
CLCT 7	6. Cross bar crosses staff but most of bar is on left side	874	0.63158
CLCT 8	7. Cross bar crosses staff but most of bar is on right side	874	0.6968
CLCT 9	8. Cross bar is approximately straight	874	0.92563
CLCT 10	9. Crossbar is connected to subsequent letter	874	0.34668
CLCT 11	10. Internal t not connected to both previous and subsequent letter	874	0.11785
CLCT 12	11. Initial or internal letter connected to subsequent letter	874	0.95881
CLCU 1	Character not present		
CLCU 2	1. Left side of letter is retrace (open or closed)	869	0.94476
CLCU 3	2. Peaks are approximately same height	869	0.97468
CLCU 4	3. Initial or internal letter connected to subsequent letter	869	0.96778
CLCU 5	4. Initial or internal letter not connected to subsequent letter	869	0.17261
CLCU 6	a. There is no down stroke on right side	869	0.04718
CLCU 7	b. Terminal stroke is approximately straight down stroke	869	0.12543
CLCV 1	Character not present		
CLCV 2	1. Internal or terminal letter connected to previous lower case letter	862	0.9652
CLCV 3	2. Initial stroke does not begin at top of staff	862	0.69722
CLCV 4	3. Left peak is loop	862	0.05336
CLCV 5	4. Internal v not connected to both previous and subsequent letter	862	0.07657
CLCV 6	5. Initial or internal letter not connected to subsequent letter	862	0.17053
CLCV 7	a. Stroke terminates at top of right upstroke	862	0.14153
CLCV 8	b. Terminal loop	862	0.00812
CLCV 9	c. Terminal angular change of direction	862	0.03132
CLCV 10	d. Other	862	0.02204
CLCW 1	Character not present		
CLCW 2	1. Internal or terminal letter connected to previous lower case letter	870	0.95862
CLCW 3	2. Initial stroke begins at top of staff	870	0.58851
CLCW 4	3. Left peak is angular movement or retrace	870	0.67816
CLCW 5	4. Left peak is loop	870	0.14713
CLCW 6	5. Left undercurve is clearly rounded (u-shaped)	870	0.95517
CLCW 7	6. Middle peak is clearly taller than both left peak and right peak	870	0.10575
CLCW 8	7. Middle peak is clearly taller than right peak but not left peak	870	0.14828
CLCW 9	8. Stroke terminates at top of right upstroke	870	0.5046
CLCW 10	9. Right peak enclosed loop	870	0.14368
CLCW 11	10. Top of right side is other	870	0.47011
CLCW 12	11. Terminal curve	870	0.13333
CLCX 1	Character not present		
CLCX 2	1. Internal or terminal letter connected to previous lower case letter	860	0.91279
CLCX 3	2. Cross strokes are connected	860	0.11512
CLCX 4	3. Cross strokes are not connected	860	0.94767
CLCX 5	4. Both crossbars are approximately straight	860	0.46977
CLCX 6	5. Internal x not connected to both previous and subsequent letter	860	0.11744
CLCX 7	6. Initial or internal letter connected to subsequent letter	860	0.76512
CLCY 1	Character not present		
CLCY 2	1. Internal or terminal letter connected to previous lower case letter	871	0.96441
CLCY 3	2. Initial stroke begins at top of bowl	871	0.68542
CLCY 4	3. Terminus is clearly sloped downward	871	0.20207
CLCY 5	4. Peaks are approximately same height	871	0.88978
CLCY 6	5. Initial or internal letter connected to subsequent letter	870	0.92759
CLCZ 1	Character not present		

CLCZ 2	1. Printed format	845	0.22012
CLCZ 3	2. Descender is curved counterclockwise	845	0.0355
CLCZ 4	3. Terminus is clearly sloped upward	845	0.5858
CLCZ 5	4. Terminus is sloped approximately horizontal	845	0.1787
CUCA 1	Character not present		
CUCA 2	1. Printed Form	855	0.39181
CUCA 3	2. Initial stroke begins to left of staff	855	0.40117
CUCA 4	3. Initial stroke with clearly clockwise curve	855	0.00585
CUCA 5	4. Initial stroke with open or closed retrace	855	0.07018
CUCA 6	5. Initial stroke with enclosed loop	855	0.01053
CUCA 7	6. Initial stroke is other	855	0.35205
CUCA 8	7. Loop is clearly wider than tall	855	0.00819
CUCA 9	8. Loop terminates and staff is a separate stroke	855	0.00351
CUCA 10	9. Terminal stroke is clearly curved counterclockwise	855	0.17895
CUCA 11	10. Terminal stroke has more than one curve	856	0.01051
CUCA 12	11. Not connected to subsequent letter	856	0.22897
CUCB 1	Character not present		
CUCB 2	1. Extraneous stroke	854	0.25995
CUCB 3	2. Initial loop clearly clockwise	854	0.01756
CUCB 4	3. Initial stroke at or near base	854	0.2178
CUCB 5	4. Upper loop height approximately equivalent to top of staff	854	0.34778
CUCB 6	5. Upper loop height clearly lower than top of staff	854	0.04684
CUCB 7	6. Upper loop with clearly angular movement	854	0.18501
CUCB 8	7. Width to height ratio imbalance of upper loop is not obvious	854	0.46838
CUCB 9	8. Buckle clearly ends to right of staff	854	0.46019
CUCB 10	9. Buckle ends approximately on staff	854	0.62295
CUCB 11	10. Terminal stroke clearly curves counterclockwise	854	0.0082
CUCB 12	11. Terminal stroke ends approximately vertical downward	854	0.03864
CUCB 13	12. Terminal stroke ends clearly downward from horizontal	854	0.12998
CUCC 1	Character not present		
CUCC 2	1. Main body clearly wider than tall	852	0.12676
CUCC 3	2. Connected to subsequent letter	852	0.94484
CUCD 1	Character not present		
CUCD 3	2. Flag initial stroke, curved	809	0.05562
CUCD 4	3. Other form of initial stroke	809	0.21508
CUCD 5	4. Lower left loop present	809	0.3115
CUCD 6	5. Lower left portion is counterclockwise movements to inside of character	809	0.03337
CUCD 7	6. Bottom of body clearly lower than bottom of staff	809	0.04697
CUCD 8	7. Bottom of body contains clearly defined angular movement	809	0.01731
CUCD 9	8. Bottom of body is other shape	809	0.54512
CUCD 10	9. Terminal stroke extends clearly to left of top of staff	809	0.12237
CUCD 11	10. Terminal stroke clearly curves counterclockwise not touching staff	809	0.12855
CUCD 12	11. Terminal stroke clearly curves counterclockwise touching staff	809	0.26452
CUCE 1	Character not present		
CUCE 2	1. Printed form	801	0.14357
CUCE 3	2 Lower arc clearly narrower than upper arc	801	0.16729
CUCE 4	3. Rounded and angular arcs in one character	801	0.07116
CUCE 5	a. Angular in top arc	801	0.04494
CUCE 6	b. Angular in bottom arc	801	0.03121
CUCE 7	4. Initial loop	801	0.15855
CUCE 8	5. Flag stroke clearly curved	801	0.01498
CUCE 9	6. Buckle clearly extends to right of right edge of arc (either arc)	801	0.08365
CUCE 10	7. Buckle pointed clearly downward	801	0.03121
CUCE 11	8. Buckle not straight	801	0.04744
CUCE 12	9. Lower arc terminate approximately horizontal	801	0.35206

CUCE 13	10. Lower arc terminates with enclosed loop	801	0.03371
CUCF 1	Character not present		
CUCF 2	1. Three stroke	765	0.39216
CUCF 3	2. Does not have a cap	765	0.01569
CUCF 4	3. Initial stroke of cap is enclosed loop	765	0.0915
CUCF 5	4. Cap is clearly curved (undercurve)	765	0.24575
CUCF 6	5. Cap has other shape	765	0.00523
CUCF 7	6. Cap is connected to staff	765	0.26797
CUCF 8	7. Connection is closed retrace	765	0.14248
CUCF 9	8. Connection is clearly defined angular movement	765	0.21699
CUCF 10	9. Connection is rounded	765	0.06536
CUCF 11	10. Bottom portion contains enclosed loop	765	0.21307
CUCG 1	Character not present		
CUCG 2	1. Printed Form	846	0.38061
CUCG 3	2. Cursive Form	846	0.59811
CUCG 4	a. Initial stroke clearly begins below bottom of body	846	0.1513
CUCG 5	b. Initial stroke is clearly curved clockwise	846	0.00473
CUCG 6	c. Staff is clearly curved	846	0.46809
CUCG 7	e. Stroke between upper left and upper right portion is clearly curved	846	0.53901
CUCG 8	f. Stroke between upper left and upper right has clearly defined angular movement	846	0.01182
CUCG 9	g. Stroke between upper left and upper right is approximately straight	846	0.02128
CUCG 10	h. Upper left peak clearly lower than upper right	846	0.17967
CUCG 11	i. Upper right is clockwise curve	846	0.01773
CUCG 12	j. Down stroke from upper right is clearly curved	846	0.26123
CUCG 13	k. Down stroke from upper right is approximately straight	846	0.17494
CUCG 14	l. Down stroke from upper right is double curve	846	0.11466
CUCG 15	m. Down stroke from upper right is other	846	0.01064
CUCG 16	n. Terminal portion is clearly clockwise curve not enclosed loop	846	0.0792
CUCG 17	o. Terminal portion is clearly counterclockwise curve	846	0.03428
CUCG 18	p. Terminal portion is extension of down stroke	846	0.00827
CUCG 19	q. Terminal portion is approximately straight	846	0.04137
CUCG 20	r. Connected to subsequent letter	846	0.40426
CUCG 21	s. Not connected to subsequent letter	846	0.22695
CUCH 1	Character not present		
CUCH 2	1. Three stroke	824	0.49272
CUCH 3	2. Two stroke	824	0.40655
CUCH 4	3. One stroke	824	0.05097
CUCH 5	4. Extraneous initial stroke present	824	0.39199
CUCH 6	a. Initial stroke is enclosed loop	824	0.05825
CUCH 7	b. Initial stroke is clearly counterclockwise curved stroke	824	0.08495
CUCH 8	c. Initial stroke is clearly clockwise curved stroke	824	0.17233
CUCH 9	d. Initial stroke is approximately straight stroke	824	0.09466
CUCH 10	5. Left staff is approximately straight	824	0.80218
CUCH 11	6. Bottom of left staff is connected to top of right staff	824	0.0449
CUCH 12	a. Stroke is approximately straight	824	0.03519
CUCH 13	b. Stroke is clearly curved clockwise	824	0.00485
CUCH 14	c. Stroke is clearly curved counterclockwise	824	0.00728
CUCH 15	7. Bottom of left staff is connected to crossbar	824	0.03155
CUCH 16	8. Bottom of left staff is clearly lower than bottom of right staff	824	0.48422
CUCH 17	9. Top of right staff is clearly higher than top of left staff	824	0.48301
CUCH 18	10. Right staff is approximately straight	824	0.74272
CUCH 19	11. Right staff is clearly curved clockwise	824	0.01942
CUCH 20	12. Right staff is clearly curved counterclockwise	824	0.11408
CUCH 21	13. Bottom of right staff is enclosed loop	824	0.06917
CUCH 22	a. Clockwise curve	824	0.03155

CUCH 23	b. Other	824	0.03519
CUCH 24	14. Bottom of right staff is not connected to crossbar	824	0.5182
CUCH 25	15. Crossbar is clearly curved clockwise	824	0.03519
CUCH 26	16. Crossbar crosses through left staff	824	0.6068
CUCH 27	17. Crossbar does not touch left staff	824	0.10801
CUCH 28	18. Crossbar crosses right staff	824	0.76335
CUCH 29	19. Crossbar does not touch right staff	824	0.04248
CUCH 30	20. Connected to subsequent letter	824	0.59345
CUCH 31	21. Not connected to subsequent letter	824	0.35922
CUCI 1	Character not present		
CUCI 2	1. San Serif Form	845	0.0568
CUCI 3	2. Printed Form with Serif(s)	845	0.20592
CUCI 4	3. Cursive Form	845	0.71124
CUCI 5	a. Staff is closed retrace	845	0.07692
CUCI 6	b. Base contains clockwise curve or loop note arrow direction	845	0.11006
CUCI 7	c. Base contains counterclockwise curve or loop note arrow direction	845	0.04024
CUCI 8	d. Base is clearly longer than staff	845	0.14083
CUCI 9	4. Connected to subsequent letter	845	0.00118
CUCJ 1	Character not present		
CUCJ 2	1. Initial stroke at or near top of letter	842	0.22684
CUCJ 3	2. “3” format	842	0.00356
CUCJ 4	3. Top portion looped	842	0.65796
CUCJ 5	4. Top portion retrace (open or closed)	842	0.0867
CUCJ 6	5. Bottom portion clockwise looped	842	0.74109
CUCJ 7	6. Bottom portion retrace (open or closed)	842	0.05463
CUCJ 8	7. Bottom portion approximately straight stroke	842	0.05463
CUCJ 9	8. Top portion and bottom portion are both enclosed loops	842	0.51663
CUCJ 10	9. Lower loop clearly wider than upper loop	842	0.12589
CUCJ 11	10. Connected to subsequent letter	842	0.73872
CUCK 1	Character not present		
CUCK 2	1. Initial stroke is curved counterclockwise	829	0.10012
CUCK 3	2. Staff clearly curved counterclockwise	829	0.23402
CUCK 4	3. Staff not connected to diagonal stroke arm or leg	829	0.74789
CUCL 1	Character not present		
CUCL 2	1. Printed Form	836	0.21531
CUCL 3	2. Modified printed form (additional stroke)	836	0.06699
CUCL 4	a. Additional stroke is initial stroke	836	0.0323
CUCL 5	b. Additional stroke is terminal stroke	836	0.01555
CUCL 6	3. Initial stroke is clearly clockwise movement	836	0.01914
CUCL 7	4. Initial portion of character loop (open or closed)	836	0.02871
CUCL 8	5. Initial portion of character retrace (open or closed)	836	0.03708
CUCL 9	6. Initial portion of character other	836	0.58612
CUCL 10	7. Bottom portion of character extends approximately the same amount to left of staff than upper portion	836	0.29187
CUCL 11	8. Bottom portion clearly does not extend as far left of staff than upper portion	836	0.09091
CUCL 12	9. Bottom left portion consists of an enclosed counterclockwise loop	836	0.01435
CUCL 13	10. Bottom left portion is other	836	0.34809
CUCL 14	11. Bottom portion terminates with clearly clockwise curve	836	0.3445
CUCL 15	12. Bottom portion terminates clearly to left of staff	836	0.00478
CUCL 16	13. Not connected to subsequent letter	836	0.30144
CUCM 1	Character not present		
CUCM 2	1. Counterclockwise curving initial stroke	845	0.14201
CUCM 3	2. Extraneous straight initial stroke	845	0.2426
CUCM 4	3. Initial stroke begins on staff	845	0.24734
CUCM 5	4. Upward stroke to first overcurve is retrace (open or closed)	845	0.61538

CUCM 6	5. Upward stroke to first overcurve is clearly counterclockwise curve (no angular point)	845	0.01183
CUCM 7	6. Left overcurve is clearly taller than right overcurve	845	0.42012
CUCM 8	7. Not connected to subsequent letter	845	0.28757
CUCN 1	Character not present		
CUCN 2	1. Printed Form	851	0.34195
CUCN 3	2. Cursive Form	851	0.63807
CUCN 4	a. Cursive form - design as shown	851	0.01998
CUCN 5	b. Counterclockwise curving initial stroke	851	0.11986
CUCN 6	c. Upward stroke to overcurve is retrace (open or closed)	851	0.40188
CUCN 7	d. Upward stroke to first overcurve is counterclockwise curve (no angular point)	851	0.00588
CUCN 8	e. Right downstroke is clockwise curve	851	0.05993
CUCN 9	f. Right downstroke is counterclockwise curve	851	0.13984
CUCN 10	g. Right downstroke has multiple curves	851	0.23032
CUCN 11	h. Connected to subsequent letter	851	0.47709
CUCO 1	Character not present		
CUCO 2	1. Initial stroke begins inside loop	856	0.49533
CUCO 3	2. Initial stroke begins outside loop	856	0.15304
CUCO 4	3. Initial stroke begins approximately on the loop	856	0.46028
CUCO 5	4. Terminus is clearly downward	856	0.18808
CUCO 6	5. Terminus is within loop	856	0.30491
CUCO 7	6. Terminal stroke is curved clockwise	856	0.00584
CUCO 8	7. Terminal stroke is curved counterclockwise	856	0.58879
CUCO 9	a. Terminus is within loop	856	0.11916
CUCO 10	b. Terminus is approximately on the loop	856	0.1507
CUCO 11	8. Not connected to subsequent letter	856	0.70794
CUCP 1	Character not present		
CUCP 2	1. 2 Stroke Design	841	0.35791
CUCP 3	2. Initial stroke is clockwise curve	841	0.05232
CUCP 4	3. Peak of staff is clearly higher than loop	841	0.07967
CUCP 5	4. Terminal stroke does not reach staff	841	0.1522
CUCP 6	5. Terminal stroke crosses staff	841	0.54697
CUCP 7	6. Terminal stroke is angular movements (closed or open retrace or angular change of direction)	841	0.09394
CUCP 8	7. Connected to subsequent letter.	841	0.12128
CUCQ 1	Character not present		
CUCQ 2	1. “2” design	848	0.11321
CUCQ 3	2. Connected to subsequent letter	848	0.3809
CUCQ 4	3. Not connected to subsequent letter	848	0.60731
CUCR 1	Character not present		
CUCR 2	1. Initial stroke is clockwise curve	855	0.0807
CUCR 3	2. Initial stroke is retrace (open or closed)	855	0.45731
CUCR 4	3. Initial stroke begins at bottom of staff	855	0.21404
CUCR 5	4. Staff disconnected	855	0.14503
CUCR 6	5. Buckle is loop	855	0.2386
CUCR 7	6. Buckle is retrace (open or closed) or angular change of direction	855	0.73918
CUCR 8	7. Not connected to subsequent letter	855	0.36491
CUCS 1	Character not present		
CUCS 2	1. Printed Form	854	0.36885
CUCS 3	2. Top loop and lower loop are approximately the same height	854	0.3466
CUCS 4	3. Terminal portion does not reach initial upstroke	854	0.08314
CUCS 5	4. Terminal portion touches initial upstroke	854	0.44028
CUCT 1	Character not present		
CUCT 2	1. Cursive Form – One Stroke	771	0.15824
CUCT 3	2. Initial stroke is enclosed loop	771	0.13878
CUCT 4	3. Initial stroke is clearly counterclockwise curved	771	0.18418

CUCT 5	4. Initial stroke is approximately straight	771	0.34112
CUCT 6	5. Non-initial stroke end of cap has enclosed loop	771	0.06615
CUCT 7	6. Non-initial stroke end of cap has retrace or angular change of direction (open or closed)	771	0.26848
CUCT 8	7. Non-initial stroke end of cap has rounded change of direction	771	0.0441
CUCT 9	8. Letter ends at base of staff	771	0.32944
CUCT 10	9. Bottom of staff has clearly defined angular change of direction	771	0.05447
CUCT 11	10. Bottom portion contains clockwise curve	771	0.21141
CUCT 12	11. Bottom portion contains clockwise enclosed loop	771	0.12322
CUCT 13	12. Bottom portion contains counterclockwise curve or enclosed loop	771	0.03502
CUCT 14	13. Bottom portion contains clearly defined angular change of direction or retrace (open or closed)	771	0.16861
CUCT 15	14. Disconnected cap is approximately straight	770	0.24805
CUCU 1	Character not present		
CUCU 2	1. Initial stroke is counterclockwise stroke	842	0.07245
CUCU 3	2. Initial stroke is approximately straight extraneous stroke	842	0.1829
CUCU 4	3. Initial stroke is non-extraneous beginning of left side down stroke	842	0.55701
CUCU 5	4. Left side is clearly bowed clockwise	842	0.09976
CUCU 6	5. Right side contains open loop	842	0.22328
CUCU 7	6. Right side contains rounded peak	842	0.08076
CUCU 8	7. Not connected to subsequent letter	842	0.30523
CUCV 1	Character not present		
CUCV 2	1. Initial stroke is enclosed loop	842	0.06651
CUCV 3	2. Initial stroke is clockwise curve (under 360 degrees)	842	0.33729
CUCV 4	3. Bottom of letter contains clearly defined angular change of direction or retrace (open or closed)	842	0.4133
CUCV 5	4. Bottom of letter is rounded	842	0.54394
CUCV 6	5. Terminal stroke is loop (over 360 degrees)	842	0.00713
CUCV 7	6. Right peak is clearly higher than left peak	842	0.57363
CUCW 1	Character not present		
CUCW 2	1. Initial stroke is counterclockwise curve/loop	862	0.04408
CUCW 3	2. Initial stroke is “3” design	862	0.00928
CUCW 4	3. Left bowl contains disconnect	862	0.0116
CUCW 5	4. Middle peak is enclosed loop	862	0.09397
CUCW 6	5. Middle peak is disconnect	862	0.00696
CUCW 7	6. Middle peak is clearly taller than left peak	862	0.06381
CUCW 8	7. Middle peak is clearly shorter than left peak	862	0.85151
CUCW 9	8. Middle peak is approximately the same height as left peak	862	0.20418
CUCW 10	9. Down stroke of middle peak is clearly curved clockwise	862	0.06265
CUCW 11	10. Bottom of right bowl is disconnect	862	0.00464
CUCW 12	11. Bottom of left bowl clearly goes lower than bottom of right bowl	862	0.45592
CUCW 13	12. Upstroke of right bowl is clearly curved counterclockwise	862	0.62413
CUCW 14	13. Upstroke of right bowl is clearly curved clockwise	862	0.11485
CUCW 15	14. Terminal stroke enclosed loop	862	0.04524
CUCW 16	15. Not connected to subsequent letter	862	0.82251
CUCX 1	Character not present		
CUCX 2	1. One stroke	843	0.08185
CUCX 3	2. Initial stroke approximately straight	843	0.45196
CUCX 4	3. Initial stroke enclosed loop	843	0.06287
CUCX 5	4. Initial stroke clockwise curve, not enclosed loop	843	0.30605
CUCX 6	5. Initial stroke counterclockwise curve, not enclosed loop	843	0.04626
CUCX 7	6. Connection is on right side	843	0.04508
CUCX 8	7. Connection is on left side	843	0.03321
CUCX 9	8. Right peak is clearly higher than left peak	843	0.33452
CUCX 10	9. Left peak is clearly higher than right peak	843	0.26335
CUCX 11	10. Left and Right bottoms are approximately same plane	843	0.34875

CUCX 12	11. Left bottom is clearly lower than right bottom	843	0.58244
CUCX 13	12. Left and Right bottoms are approximately same plane	843	0.25979
CUCX 14	13. Intersection of lines is clearly above the midpoint	843	0.12574
CUCX 15	14. Intersection of lines is clearly below the midpoint	843	0.36892
CUCX 16	15. Intersection of lines is approximately at midpoint	843	0.46738
CUCX 17	16. Top angle is clearly greater than 90 degrees	843	0.01423
CUCY 1	Character not present		
CUCY 2	1. Printed Form	820	0.17439
CUCY 3	2. Initial stroke is clearly counterclockwise curve/loop	820	0.06585
CUCY 4	3. Right peak is clearly taller than left peak	820	0.34146
CUCY 5	4. Left peak is clearly taller than right peak	820	0.32805
CUCY 6	5. Descender is approximately straight and terminates at end down stroke	820	0.15366
CUCY 7	6. Descender contains counterclockwise curve/loop	820	0.00854
CUCY 8	7. Descender is clearly clockwise curve of approximately 90 degrees	820	0.01098
CUCY 9	8. Descender is clockwise curve of over 90 degrees but does not intersect down stroke	820	0.06829
CUCY 10	9. Descender intersects upper bowl	820	0.08171
CUCZ 1	Character not present		
CUCZ 2	1. Printed Form	855	0.55556
CUCZ 3	b. With crossbar	855	0.09357
CUCZ 4	2. Cursive Form	855	0.42456
CUCZ 5	a. Initial stroke is clearly curved counterclockwise	855	0.01053
CUCZ 6	b. Buckle clearly does not go as far left as upper portion	855	0.21053
CUCZ 7	c. Buckle clearly goes farther to the left than upper portion	855	0.03275
CUCZ 8	d. Buckle and upper portion are approximately the same extent left	855	0.16374
CUCZ 9	e. Buckle is loop	854	0.05035
CUCZ 10	f. Buckle is retrace (open or closed)	854	0.33607
CUCZ 11	g. Buckle is curved/missing	854	0.01288
CUCZ 12	h. Descender is approximately straight	854	0.02108
CUCZ 13	i. Descender is clearly counterclockwise curve/loop	854	0.00468
CUCZ 14	j. Descender is enclosed loop	854	0.38056

Table 16 – 435 cursive features listed, description of letter, "count" (population) of that letter (used for determining standard deviation and confidence limits), and the frequency occurrence proportion for each feature. The database with illustrations may provide assistance in understanding the specifics of each feature.

Table 17 includes the frequency occurrence proportions for each of the selected hand printed characteristics. The feature terms are abbreviated. "PLC" represents "Printed Lower Case" thus "PLCF" represents "Printed Lower Case "f" and "PLCR" represents "Printed Lower Case "r". Similarly, "PUC" represents "Printed Upper Case. Therefore, "PUCA" represents "Printed Upper Case "A".

FEATURE	DESCRIPTION	COUNT	FREQUENCY
PLCA 1	Character not present		
PLCA 2	1. Cap or initial stroke is to the left of the peak of the staff	697	0.79627
PLCA 3	2. The staff is an open loop	697	0.26686
PLCA 4	3. The staff is a retrace (open and closed)	697	0.91535
PLCB 1	Character not present		
PLCB 2	1. Initial stroke begins at top of staff	746	0.95845
PLCB 3	2. Staff is approximately straight	746	0.91421
PLCB 4	3. Staff is clearly bowed or curved	746	0.30563
PLCB 5	4. Initial portion of bottom loop is angular change of direction or retrace (open or	746	0.82976

	closed) with staff		
PLCB 6	5. Bottom loop is counterclockwise ("6" design)	746	0.06166
PLCB 7	6. Design is other than standard design	746	0.0496
PLCD 1	Character not present		
PLCD 2	1. Initial stroke is curved clockwise	759	0.05534
PLCD 3	2. Staff is clearly bowed	759	0.22793
PLCD 4	3. Connection is enclosed loop	759	0.08169
PLCD 5	4. Staff and loop are not connected (2 stroke form)	759	0.18182
PLCD 6	5. Loop is clearly taller than wide	759	0.14493
PLCD 7	6. Initial stroke begins lower loop	759	0.49275
PLCD 8	7. Loop is not open	759	0.70224
PLCD 9	8. Width to height ratio imbalance is not obvious	759	0.76285
PLCD 10	9. Staff is primarily closed retrace	759	0.59289
PLCD 11	10. Overall design is other	759	0.07905
PLCE 1	Character not present		
PLCE 2	1. Letter is enclosed loop	739	0.97835
PLCF 1	Character not present		
PLCF 2	1. Bottom portion of staff is approximately straight	717	0.92608
PLCG 1	Character not present		
PLCG 2	1. Upper loop is clearly clockwise movement	717	0.18828
PLCG 3	2. Width to height ratio imbalance is not obvious	717	0.82566
PLCG 4	3. Top of staff is open loop	717	0.17852
PLCG 5	4. Lower extender is enclosed loop, triangulation or other similar design	717	0.2636
PLCG 6	a. Lower loop/design intersects upper loop	717	0.03487
PLCG 7	b. Lower loop/design does not intersect upper loop	717	0.2357
PLCH 1	Character not present		
PLCH 2	1. Initial stroke begins at top of staff	746	0.99062
PLCH 3	2. Staff is approximately straight	746	0.94906
PLCH 4	3. Connection from staff to overcurve is open loop	746	0.05094
PLCH 5	4. Connection from staff to overcurve is retrace (open or closed)	746	0.92761
PLCH 6	5. Staff and overcurve are not connected	746	0.03485
PLCH 7	6. Overcurve is rounded	746	0.91555
PLCH 8	7. Bottom of staff and right side of overcurve are approximately on a level horizontal plane	746	0.92761
PLCJ 1	Character not present		
PLCJ 2	1. Initial stroke of staff is curved	752	0.11303
PLCJ 3	2. Initial stroke of staff is approximately straight	752	0.9508
PLCJ 4	3. Bottom of staff clearly curves more than 180 degrees	752	0.16755
PLCJ 5	4. Dot is present	752	0.88697
PLCJ 6	a. Clearly diagonal as shown	752	0.13298
PLCJ 7	b. Circle	752	0.02261
PLCK 1	Character not present		
PLCK 2	1. Initial stroke begins at top of staff	759	0.99605
PLCK 3	2. Bottom right stroke is not approximately straight	759	0.83267
PLCL 1	Character not present		
PLCL 2	1. General design is single approximately vertical stroke	735	0.97551
PLCM 1	Character not present		
PLCM 2	1. Initial stroke begins at top of staff	752	0.95612
PLCM 3	2. Presence of extraneous initial stroke	752	0.03856
PLCM 4	3. Left leg contains retrace (open or closed)	752	0.82846
PLCM 5	4. Left leg is loop	752	0.03856
PLCM 6	5. Left overcurve has clearly defined point	752	0.48005
PLCM 7	6. Middle leg is enclosed loop	752	0.03989
PLCN 1	Character not present		
PLCN 2	1. Initial stroke begins at top of staff	717	0.95537

PLCN 3	2. Left leg contains retrace (open or closed)	717	0.841
PLCN 4	3. Down stroke after overcurve is approximately straight for majority of length	717	0.76011
PLCO 1	Character not present		
PLCO 2	1. Loop is closed	768	1
PLCP 1	Character not present		
PLCP 2	1. Staff is loop	751	0.31158
PLCP 3	2. Staff is retrace (open or closed)	751	0.80692
PLCP 4	3. Width to height ratio imbalance is not obvious	751	0.87883
PLCQ 1	Character not present		
PLCQ 2	b. Lower loop/design does not intersect upper loop	670	0.58507
PLCR 1	Character not present		
PLCR 2	1. Initial stroke begins at top of staff	713	0.96213
PLCR 3	2. Staff is approximately straight	713	0.96634
PLCR 4	3. Staff and cap are disconnected	713	0.03927
PLCS 1	Character not present		
PLCS 2	1. Left side of top bowl does not have predominantly defined angular movement	737	0.94301
PLCS 3	2. Slope is predominantly downward right to left	737	0.59837
PLCT 1	Character not present		
PLCT 2	1. Initial stroke begins at top of staff	757	0.98943
PLCT 3	2. Staff is approximately straight	757	0.96037
PLCT 4	3. Crossbar is present	757	0.97226
PLCT 5	4. Crossbar bisects staff	757	0.93659
PLCT 6	5. Crossbar is not present	757	0.0251
PLCT 7	6. Crossbar connected to subsequent letter	757	0.60766
PLCU 1	Character not present		
PLCU 2	1. Initial stroke begins at top of staff	769	0.9909
PLCU 3	2. Peaks are approximately same height	769	0.93498
PLCU 4	3. Right side terminates at apex	769	0.39272
PLCU 5	4. Right side does not terminate at apex	769	0.69311
PLCU 6	a. Right side of letter is retrace (open or closed)	769	0.62809
PLCV 1	Character not present		
PLCV 2	1. Initial stroke begins at top of staff	768	0.98698
PLCV 3	2. One stroke	768	0.8151
PLCW 1	Character not present		
PLCW 2	1. Initial stroke begins at top of staff	764	0.99084
PLCW 3	2. Bottom of left valley is clearly rounded (u-shaped)	764	0.87827
PLCW 4	3. Middle peak is taller than both left peak and right peak	764	0.07199
PLCW 5	4. Middle peak is loop	764	0.10471
PLCW 6	5. Four stroke	764	0.01571
PLCX 1	Character not present		
PLCX 2	1. Cross strokes are connected	614	0.10261
PLCY 1	Character not present		
PLCY 2	1. Initial stroke begins at top left	761	0.94612
PLCY 3	2. One stroke	761	0.5138
PLCY 4	3. One stroke is approximately straight of 2 stroke design	761	0.63863
PLCY 5	4. Descender is enclosed loop, triangulation or other design	761	0.16426
PLCY 6	a. Lower loop/design intersects upper loop	761	0.01971
PLCY 7	b. Lower loop/design does not intersect upper loop	761	0.16032
PLCZ 1	Character not present		
PLCZ 2	1. One stroke	762	0.90551
PLCZ 3	2. Two stroke	762	0.021
PLCZ 4	3. Three stroke	762	0.00131
PLCZ 5	a. Top stroke and angular stroke cross	762	0
PLCZ 6	b. Top stroke and angular stroke do not touch	762	0
PLCZ 7	c. Angular stroke and bottom stroke cross	762	0

PLCZ 8	d. Angular stroke and bottom stroke do not touch	762	0
PLCZ 9	4. Crossbar present	762	0.17192
PUCA 1	Character not present		
PUCA 2	1. Printed Format	821	0.97686
PUCA 3	2. Terminal counterclockwise curve into crossing	821	0.04507
PUCA 4	a. Other	821	0.01949
PUCA 5	3. Left side stroke approximately straight	821	0.85627
PUCA 6	4. Left side stroke with double curve (or more)	821	0.01705
PUCA 7	5. Right side stroke with double curve (or more)	821	0.02071
PUCA 8	6. Left side stroke primarily past vertical	821	0.05725
PUCA 9	7. Left leg clearly longer than right leg	821	0.54933
PUCA 10	8. Right leg clearly longer than left leg	821	0.33861
PUCA 11	9. Staff is enclosed loop	821	0.0609
PUCA 12	10. Initial flag stroke	821	0.00853
PUCA 13	11. Enclosed loop in initial stroke	821	0.00122
PUCA 14	12. Flag clearly starts higher than shoulder	821	0.00365
PUCA 15	13. Flag clearly starts lower than shoulder	821	0.00122
PUCA 16	14. Indeterminate relative horizontal heights	821	0.00122
PUCA 17	15. Flag clearly starts below peak of letter	821	0.00244
PUCA 18	16. Flag starts approximately level to peak of letter	821	0
PUCA 19	17. Crossbar is clearly overcurve	821	0.08526
PUCA 20	18. Crossbar never touches right side stroke	821	0.05725
PUCA 21	19. Cursive Format	821	0.05968
PUCB 1	Character not present		
PUCB 2	1. Initial stroke at top of staff	816	0.90686
PUCB 3	2. Extraneous initial stroke present	816	0.00858
PUCB 4	3. Middle connector ends to right of staff	816	0.54289
PUCB 5	4. Middle connector ends on staff	816	0.59069
PUCB 6	5. Terminal stroke clearly curves clockwise	816	0.50613
PUCB 7	6. Terminal stroke enclosed loop	816	0.01471
PUCB 8	7. Terminal stroke ends approximately horizontal	816	0.3076
PUCC 1	Character not present		
PUCC 2	1. Initial stroke is retrace (open or closed)	815	0.24908
PUCC 3	2. Initial stroke clearly curves clockwise	815	0.00123
PUCC 4	3. Initial stroke at end of curve (no extraneous stroke)	815	0.79755
PUCC 5	4. Main body clearly taller than wide	815	0.4589
PUCC 6	5. Main body contains clearly angular movement in bottom half	815	0.02577
PUCC 7	6 .Terminal stroke counterclockwise enclosed loop	815	0.00245
PUCC 8	7. Terminal stroke curves clockwise	815	0.01104
PUCD 1	Character not present		
PUCD 2	1. Initial stroke is down stroke of staff	812	0.95443
PUCD 3	2. Flag initial stroke, curved	812	0.00123
PUCD 4	3. Flag initial stroke, straight	812	0.00616
PUCD 5	4. Other form of initial stroke	812	0.03818
PUCD 6	5. Initial stroke of loop does not touch staff or reach vertical plane of staff	812	0.03695
PUCD 7	6. Staff clearly not straight	812	0.02833
PUCD 8	7. Staff approximately straight	811	0.90752
PUCD 9	8. Staff connected to loop	811	0.1307
PUCD 10	9. Loop has clearly defined angular movement	811	0.02836
PUCE 1	Character not present		
PUCE 2	1. One stroke	814	0.01843
PUCE 3	2. Four stroke	814	0.42015
PUCE 4	3. Greek epsilon form	814	0.14005
PUCE 5	4. C with middle horizontal stroke form	814	0.02948
PUCE 6	5. Staff is retrace (open or closed)	814	0.01966

PUCE 7	6. Staff contains enclosed loop.	814	0.00123
PUCE 8	7. Top horizontal stroke is approximately straight	814	0.71253
PUCE 9	8. Middle horizontal stroke is approximately horizontal	814	0.63759
PUCE 10	9. Bottom horizontal stroke is approximately horizontal	814	0.48894
PUCE 11	10. Top horizontal stroke is connected to middle horizontal stroke	814	0.01351
PUCE 12	11. Middle horizontal stroke is connected to bottom horizontal stroke	814	0.02088
PUCE 13	12. Top and bottom horizontal stroke are approximately equal length	814	0.45577
PUCF 1	Character not present		
PUCF 2	1. Initial stroke begins at top of staff	795	0.79497
PUCF 3	2. Extraneous initial stroke present	795	0.01006
PUCF 4	3. Top horizontal stroke does not reach vertical plane of staff	795	0.03145
PUCF 5	4. Top horizontal stroke clearly crosses vertical plane of staff but does not touch staff	795	0.1283
PUCF 6	5. Staff connected to top horizontal stroke	795	0.36478
PUCF 7	6. Top horizontal stroke clearly crosses staff	795	0.24151
PUCF 8	7. Top horizontal stroke does not reach vertical plane of staff	795	0.01132
PUCF 9	8. Clearly downward slope	795	0.03648
PUCF 10	9. Lower horizontal stroke is approximately straight	795	0.8956
PUCF 11	a. Clearly upward stroke	795	0.43396
PUCF 12	b. Clearly downward stroke	795	0.01635
PUCF 13	c. Approximately horizontal stroke	795	0.45786
PUCF 14	10. Lower horizontal stroke is not approximately straight	795	0.05535
PUCF 15	11. Lower horizontal stroke does not touch staff	795	0.04403
PUCF 16	12. Bottom horizontal stroke is clearly longer than top horizontal stroke	795	0.14088
PUCG 1	Character not present		
PUCG 2	1. Six design	816	0.26716
PUCG 3	2. Lower case design	816	0.00735
PUCG 4	3. Semi-circle design other than "six" design	816	0.71324
PUCG 5	a. Initial stroke closed loop	816	0.01716
PUCG 6	b. Initial stroke curve counterclockwise	816	0.15686
PUCG 7	c. Initial stroke approximately straight	816	0.27451
PUCG 8	d. Initial stroke curve clockwise	816	0.00245
PUCG 9	e. Main body taller than wide	816	0.21569
PUCG 10	f. Main body wider than tall	816	0.03554
PUCG 11	g. Width to height ratio imbalance is not obvious	816	0.45833
PUCG 12	h. Main body contains clearly angular movement in top half	816	0.01961
PUCG 13	i. Main body contains clearly angular movement in bottom half	816	0.00613
PUCG 14	j. Crossbar and descender present	816	0.30637
PUCG 15	k. Crossbar only present	816	0.39338
PUCG 16	l. Descender only present	816	0.02083
PUCG 17	4. Other form	816	0.00735
PUCH 1	Character not present		
PUCH 2	1. Initial stroke is enclosed loop	807	0.00124
PUCH 3	2. Initial stroke is clearly counterclockwise curved stroke	807	0.00248
PUCH 4	3. Initial stroke is clearly clockwise curved stroke	807	0.01859
PUCH 5	4. Stroke is approximately horizontal	807	0.00372
PUCH 6	5. Stroke is clearly downward from horizontal	807	0.00248
PUCH 7	6. Initial stroke is other	807	0.88352
PUCH 8	7. Bottom of left staff is connected to crossbar	807	0.01983
PUCH 9	8. Bottom of left staff is clearly lower than bottom of right staff	807	0.45229
PUCH 10	9. Top of right staff is clearly higher than top of left staff	807	0.53779
PUCH 11	10. Angular change of direction to begin cross stroke	807	0.06691
PUCH 12	11. Clockwise curve/loop to begin cross stroke	807	0.00991
PUCH 13	12. Counterclockwise curve/loop	807	0.00991
PUCH 14	13. Other	807	0.08178

PUCH 15	14. Crossbar is approximately horizontal	807	0.6109
PUCH 16	15. Crossbar is clearly overcurve	807	0.06568
PUCH 17	16. Crossbar does not touch left staff	807	0.13631
PUCH 18	17. Crossbar does not touch right staff	807	0.03594
PUCI 1	Character not present		
PUCI 2	1. San Serif Form	810	0.38148
PUCI 3	a. Staff is approximately straight	810	0.36049
PUCI 4	b. Staff is clearly curved clockwise	810	0.00864
PUCI 5	c. Staff is clearly curved counterclockwise	810	0.00988
PUCI 6	d. Staff is clearly curved more than once	810	0.00247
PUCI 7	e. Initial stroke is approximately straight	810	0.35679
PUCI 8	f. Initial stroke is clearly curved clockwise	810	0.00617
PUCI 9	g. Initial stroke is clearly curved counterclockwise	810	0.0037
PUCI 10	h. Terminal stroke is approximately straight	810	0.34198
PUCI 11	i. Terminal stroke is clearly curved clockwise	810	0.00123
PUCI 12	j. Terminal stroke is clearly curved counterclockwise	810	0.02099
PUCI 13	2. With Serif(s)	810	0.78272
PUCI 14	a. Top serif is approximately straight	810	0.63827
PUCI 15	b. Top serif is not approximately straight	810	0.14568
PUCI 16	c. Top serif is clearly left of center	810	0.18025
PUCI 17	d. Top serif is not present	810	0.00247
PUCI 18	e. Staff is clearly curved counterclockwise	810	0.0321
PUCI 19	f. Staff is clearly curved more than once	810	0.01728
PUCI 20	g. Bottom serif is not approximately horizontal	810	0.43086
PUCI 21	h. Bottom serif is not present	810	0.00494
PUCJ 1	Character not present		
PUCJ 2	1. Cap not present	815	0.31779
PUCJ 3	2. Cap present	815	0.71902
PUCJ 4	3. Staff does not touch cap	815	0.33006
PUCK 1	Character not present		
PUCK 2	1. Initial stroke begins at top of staff	819	0.98657
PUCK 3	a. Enclosed Loop	819	0
PUCK 4	b. Clearly curved	819	0.00244
PUCK 5	c. Clockwise	819	0.00855
PUCK 6	d. Counterclockwise	819	0.00122
PUCK 7	2. Staff is approximately straight	819	0.94994
PUCK 8	3. Staff is clearly curved clockwise	819	0.07937
PUCK 9	4. Staff other	819	0.01954
PUCK 10	5. Staff connected to diagonal stroke	819	0.0525
PUCK 11	6. Buckle is disconnected	819	0.27228
PUCK 12	7. Buckle is open loop	819	0.0928
PUCK 13	8. Buckle is curved stroke	819	0.21245
PUCK 14	9. Buckle does not touch or cross staff	819	0.26862
PUCL 1	Character not present		
PUCL 2	1. No extraneous strokes or ticks at top of staff	808	0.92698
PUCL 3	2. Base is approximately horizontal	808	0.59406
PUCL 4	3. Base has clearly downward slope	808	0.05941
PUCL 5	4. Lip is present at end of base	808	0.05693
PUCM 1	Character not present		
PUCM 2	1. Counterclockwise curving initial stroke	809	0.00371
PUCM 3	2. Clockwise curving initial stroke	809	0.01978
PUCM 4	3. Initial stroke begins on staff	809	0.5513
PUCM 5	4. Upward stroke to first overcurve is retrace (open or closed)	809	0.41533
PUCM 6	5. Upward stroke to first overcurve is clearly counterclockwise curve (no angular point)	809	0.01854

PUCM 6	6. Left overcurve has clearly defined pointed angular peak	809	0.62176
PUCM 7	7. Left overcurve does not have clearly defined pointed angular peak	809	0.26823
PUCM 8	8. Middle leg is longer than left leg	809	0.01236
PUCM 9	9. Middle leg is approximately same horizontal plane as left leg	809	0.19901
PUCM 10	01. Right overcurve has clearly defined pointed angular peak	809	0.65513
PUCM 11	11. Down stroke of right overcurve is clearly clockwise curve	809	0.09394
PUCN 1	Character not present		
PUCN 2	1. Cursive Form	807	0.0347
PUCN 3	2. One stroke	807	0.63073
PUCN 4	3. Two stroke	807	0.114
PUCN 5	4. Three stroke	807	0.22677
PUCN 6	5. Initial stroke at top of staff	807	0.49814
PUCO 1	Character not present		
PUCO 2	1. Terminus is approximately horizontal	790	0.11899
PUCO 3	2. Terminus is clearly downward	790	0.27215
PUCO 4	3. Terminus is within loop	790	0.58608
PUCO 5	4. Terminal stroke is curved clockwise	790	0.00506
PUCO 6	5. Terminus is outside loop	790	0.09494
PUCO 7	6. Terminus is approximately on the loop	790	0.26962
PUCP 1	Character not present		
PUCP 2	1. Initial stroke is counterclockwise curve	736	0.00272
PUCP 3	a. Initial stroke begins clearly above the peak of the loop	736	0
PUCP 4	b. Initial stroke begins clearly below the peak of the loop	736	0.00272
PUCP 5	c. Initial stroke begins approximately at same height as peak of loop	736	0.00408
PUCP 6	2. Retrace present	736	0.40489
PUCP 7	3. Peak of staff is clearly higher than loop	736	0.01902
PUCP 8	4. Terminal stroke is clearly curved counterclockwise	736	0.01766
PUCQ 1	Character not present		
PUCQ 2	1. Initial stroke begins clearly inside loop	813	0.50677
PUCQ 3	2. Initial stroke begins clearly outside loop	813	0.16482
PUCQ 4	3. Terminal stroke of loop is curved clockwise	813	0.00369
PUCQ 5	4. Stick stroke is approximately straight	813	0.67282
PUCQ 6	5. Stick stroke is not approximately straight	813	0.31242
PUCQ 7	6. Stick stroke is connected to loop	813	0.02583
PUCR 1	Character not present		
PUCR 2	1. One stroke design	761	0.48489
PUCR 3	a. Initial stroke is counterclockwise curve	761	0.00131
PUCR 4	b. Extraneous straight initial stroke	761	0.00394
PUCR 5	2. Buckle is enclosed loop	761	0.10381
PUCS 1	Character not present		
PUCS 2	1. Slope is clearly upward right to left	811	0.03946
PUCS 3	2. Slope is approximately horizontal	811	0.18619
PUCS 4	3. Right side of lower bowl has clearly defined angular movement	811	0.12577
PUCS 5	4. Right side of lower bowl does not have clearly defined angular movement	811	0.8545
PUCS 6	5. Slope is clearly upward right to left	811	0.06782
PUCS 7	6. Lower bowl clearly is further to the right than the initial stroke	811	0.2873
PUCT 1	Character not present		
PUCT 2	1. Staff is approximately straight	820	0.98293
PUCT 3	2. Bottom of staff is connected to crossbar	820	0.02683
PUCT 4	3. Crossbar is approximately straight	820	0.91951
PUCT 5	a. Crossbar has clearly upward slope	820	0.56098
PUCT 6	b. Crossbar has clearly downward stroke	820	0.05854
PUCU 1	Character not present		
PUCU 2	1. Initial stroke is clearly clockwise stroke	812	0.02586
PUCU 3	a. Initial stroke is clearly closed loop	812	0.00123

PUCU 4	b. Initial stroke is not closed loop	812	0.01724
PUCU 5	c. Slope is downward	812	0.00246
PUCU 6	d. Slope is approximately horizontal	812	0.00616
PUCU 7	2. Initial stroke is non-extraneous beginning of left side down stroke	812	0.93473
PUCU 8	3. Left side is clearly bowed clockwise	812	0.01478
PUCU 9	4. Left side has multiple curves	812	0.00862
PUCU 10	5. Right side contains open loop	812	0.10961
PUCU 11	6. Right side contains retrace (open or closed)	812	0.35591
PUCU 12	7. Left peak is clearly higher than right peak	812	0.27217
PUCU 13	8. Right peak is higher than left peak	812	0.28325
PUCU 14	9. Peaks are approximately equal height	812	0.42118
PUCU 15	10. Terminal stroke is curved clockwise	812	0.01478
PUCU 16	11. Character terminates at top of right side (no staff)	812	0.44335
PUCV 1	Character not present		
PUCV 2	1. Initial stroke is clockwise curve (under 360 degrees)	715	0.06713
PUCV 3	2. Left stroke is clearly curved counterclockwise	715	0.33007
PUCV 4	3. Right stroke is clearly curved counterclockwise	715	0.35804
PUCV 5	4. Terminal stroke clearly curves counterclockwise	715	0.02378
PUCV 6	5. Terminal stroke is loop (over 360 degrees)	715	0.00559
PUCV 7	6. Right peak is clearly lower than left peak	715	0.18322
PUCW 1	Character not present		
PUCW 2	1. Initial stroke is “3” design	821	0.00122
PUCW 3	2. Left bowl contains disconnect	821	0.0475
PUCW 4	3. Middle peak is curved stroke	821	0
PUCW 5	4. Middle peak is disconnect	821	0.07552
PUCW 6	5. Down stroke of middle peak is clearly curved clockwise	821	0.51766
PUCW 7	6. Down stroke of middle peak has multiple curves	821	0.64677
PUCW 8	7. Bottom of right bowl is disconnect	821	0
PUCW91	8. Bottom of left bowl clearly goes lower than bottom of right bowl	821	0.00853
PUCW 10	9. Terminal stroke enclosed loop	821	0
PUCX 1	Character not present		
PUCX 2	1. One stroke	821	0.02558
PUCX 3	a. Connection is on right side	821	0.0134
PUCX 4	b. Connection is on left side	821	0.01096
PUCX 5	c. Right peak is clearly higher than left peak	821	0.01218
PUCX 6	d. Left peak is clearly higher than right peak	821	0.0134
PUCX 7	e. Left and Right bottoms are approximately same plane	821	0.00487
PUCX 8	f. Right bottom is clearly lower than left bottom	821	0.00731
PUCX 9	g. Left bottom is clearly lower than right bottom	821	0.01705
PUCX 10	h. Left and Right bottoms are approximately same plane	821	0.00609
PUCX 11	i. Intersection of lines is clearly above the midpoint	821	0.00853
PUCX 12	j. Intersection of lines is clearly below the midpoint	821	0.0134
PUCX 13	k. Intersection of lines is approximately at midpoint	821	0.00853
PUCX 14	l. Top angle is clearly greater than 90 degrees	821	0
PUCX 15	m. Top angle is clearly less than 90 degrees	821	0.00853
PUCX 16	n. Top angle is approximately 90 degrees	821	0.00853
PUCX 17	2. Two Stroke	821	0.95981
PUCX 18	a. Right peak is clearly higher than left peak	821	0.34348
PUCX 19	b. Left peak is clearly higher than right peak	821	0.23995
PUCX 20	c. Left and Right peaks approximately same height	821	0.37028
PUCX 21	d. Right bottom is clearly lower than left bottom	821	0.16322
PUCX 22	e. Left bottom is clearly lower than right bottom	821	0.49939
PUCY 1	Character not present		
PUCY 2	1. Initial stroke is extraneous stroke	744	0.01344
PUCY 3	2. Left peak is clearly taller than right peak	744	0.24597

PUCY 4	3. Right peak is clearly taller than left peak	744	0.51344
PUCY 5	4. Lower extender is enclosed loop	744	0.01747
PUCZ 1	Character not present		
PUCZ 2	1. Three stroke	677	0.04874
PUCZ 3	a. Top stroke and angular stroke cross	677	0.00886
PUCZ 4	b. Top stroke and angular stroke do not touch	677	0
PUCZ 5	c. Angular stroke and bottom stroke cross	677	0.00886
PUCZ 6	d. Angular stroke and bottom stroke do not touch	677	0
PUCZ 7	2. Two stroke	677	0.05318
PUCZ 8	3. Top angle extends further to right than bottom stroke	677	0.22009

Table 17 – 351 hand printed features listed, description of letter, "count" (population) of that letter (used for determining standard deviation and confidence limits), and the frequency occurrence proportion for each feature. The database with illustrations may provide assistance in understanding the specifics of each feature.

Cursive

Since feature presence can be thought of as an indicator variable (i.e., either 1 for present/TRUE or 0 for absent/FALSE), we can further consider each variable as having a binomial distribution. The individual samples are effectively independent (no one was copying from another participant) and we posit the overall probability of presence as an unknown parameter p_i, where i goes from 1 to 435 (the number of cursive characteristics). For calculations in this project p_i will be the calculated frequency occurrence proportions listed in Table 16. The standard deviation of the proportion estimator is given as $[p_i (1- p_i)/880]^{0.5}$. (The denominator 880 will vary slightly depending on the actual count of the letter as referenced in the third column of Table 16.) This expression attains a maximum of 0.017 for $p_i = 0.5$ and is rather less as the proportion approaches either 0 or 1.

Using standard statistical methodology the 95% confidence limits of the proportions will be a range from -2 sig to +2 sig of those proportions. "Sig" is short for "sigma" and is defined as meaning the estimated standard deviation. The use of 2 rather than 1.96 as the multiplier is owed to the limits being approximations, but excellent ones for the sample sizes. The term "95% confidence limits" is defined as meaning that there is a 95% confidence level that the frequency occurrence will fall within the calculated range that is the standard for reaching statistically valid results for an entire population in a sampling environment. By applying the standard deviation estimator expression $[p_i (1- p_i)/880]^{0.5}$ to any of the cursive frequency occurrence proportions one can easily calculate the 95% confidence limits by adding and subtracting the standard deviation times two. For example, CLCM 9 has a frequency occurrence of 0.157 and a count of 873 per Table 16. Based on these numbers the standard deviation is calculated to be $[p_i (1- p_i)/880]^{0.5}$ or in this case $[0.157(1-0.157)/873]^{0.5}$ which is 0.0125. The 95% confidence limits are then calculated as 0.157 plus or minus (0.0125 X 2). The 95% confidence limits for CLCM 9 are within the range of 0.132 and 0.182. There have been previous claims that unless the entire population is tested one cannot have statistically valid

calculations as to frequency occurrence. Those claims are incorrect and contrary to standard statistical practices.

As noted earlier in this report, Huber and Headrick (1999) describe *qualitatively* various demographic features that influence handwriting in general. With the project sample established and in conjunction with the associated demographics, we can *quantitatively* assess character features as a function of age, gender, ethnicity, education level, location of cursive training, and handedness. For each combination of character feature (435 for cursive) and demographic (6 possibilities), the association as measured by Fisher's exact test has been run. A significant association implies that the demographic variable influences the presence/absence of a feature. Table 18 summarizes the results for all 435 features across the 6 demographic variables. The demographic row total basis values vary depending on the number of unknowns for the demographic category or a sub-category deliberately not included—e.g., Native American for Ethnic or ambidextrous for Right or Left which have very small sample sizes.

p-value range	Age	Location 2nd or 3rd Grade	Gender	Education	Ethnic	Right or Left
<.0001	13.1%	11.9%	2.5%	1.4%	0.2%	0.2%
<.001	17.0%	19.0%	8.7%	3.9%	1.1%	0.2%
<.01	28.2%	26.8%	16.1%	11.2%	6.9%	1.1%
<.05	40.8%	35.3%	28.7%	21.8%	16.3%	5.3%
<.1	50.0%	42.2%	37.6%	28.4%	23.4%	9.4%
>.95	1.8%	12.8%	6.9%	7.1%	3.2%	18.1%
1	1.4%	12.8%	6.9%	7.1%	1.6%	18.1%
Demographic Row Total Basis	870	696	874	861	834	869

Table 18 - Percentage of 435 Features Having Indicated p-value Range.

Table 18 is arranged from left to right according to the strength of association (stronger to weaker). Overall, age has the greatest bearing on the number of features present with location of second/third grade training a close second. Over a quarter of the features considered exhibit an effect on the presence/absence due to age of provider or location. Gender and education also exhibit a significant signal for many of the features (many more than would be expected due to chance alone). The ethnic category (restricted to Caucasian, African American, Hispanic and Asian) also influences a number of features presence/absence. Only handedness appears to have little to do with influencing presence/absence with percentages matching those that would be observed due to chance alone.

Closeness of writers based on presence/absence of features. With the database, we examined those specimens that had no missing characteristics (521 specimens). Using cluster analysis on this group of complete entries, it was found that the four pairs of "closest" specimens in terms of presence/absence of features were original specimens (#486, #712) with 36 features that differed and pairs of original specimens (#96, #123), (#4, #124), and (#126, #1110) with each 38 features that differed. Any other two specimens have more than 38 features that are different with likely 80 to 100 features not in common. It must be emphasized that even a complete match does not suggest that writers are one and the same or that these collection of presence/absence features are sufficient to completely determine authorship. Two specimen providers could conceivably exhibit the same set of database responses. However, upon examination by an expert forensic document examiner, additional subtleties and variances within the present/absent modes would be noted to distinguish authorship. In other words, reliance solely on the degree of matching is contrary to the intent of our research and to appropriate forensic document examination methodology (see Figure 6).

Product Rule Analyses. The product rule has been defined as "the probability of concurrence of all the events is equal to the continued product of the probabilities of all the separate events" (Osborn (1929) referencing Professor Simon Newcomb). A key factor in using the product rule is that each event must be independent of the other events in order to be applicable. It is recognized as a convenient tool if in fact it applies. The database allows us to consider numerous instances to test the appropriateness of the product rule with respect to presence or absence of combinations of characteristics. Since there were 436 distinct features to analyze then (one feature pared subsequent to interdependency testing), there are a total of 94,830 possible pairs of features that could be considered. Some interesting pairs can be identified by computing all pairwise correlations to obtain the extremes, intermediate and zero correlation values. Three specific pairs are considered, as an illustration of the calculations. Table 19 reveals an actual case (CLCV6 with CLCV7) for which the Product rule does not work owing to the large correlation of 0.9. Table 20 (CLCC2 with CLCE3) covers a case where the product rule works though not perfectly (correlation of 0.2). Finally, Table 21 is an example (CUCH9 and CUCK4) in which the Product rule works nearly flawlessly (correlation = 0.0).

		Count	CLCV6		Total
			FALSE	TRUE	
CLCV7	FALSE	Observed	712	28	740
		Expected with Product Rule	613.8	126.2	
	TRUE	Observed	3	119	122
		Expected with Product Rule	101.2	20.8	
		Total	715	147	862

Table 19 - Product rule does not work (correlation of 0.9).

		Count	CLCC2		Total
			FALSE	TRUE	
CLCE3	FALSE	Observed	7	7	14
		Expected with Product Rule	0.8	13.2	
	TRUE	Observed	41	810	851
		Expected with Product Rule	47.2	803.8	
		Total	48	817	865

Table 20 - Product rule works relatively well (correlation of 0.2).

		Count	CUCH9		Total
			FALSE	TRUE	
CUCK4	FALSE	Observed	184	17	201
		Expected with Product Rule	181.8	19.2	
	TRUE	Observed	535	59	594
		Expected with Product Rule	537.2	56.8	
		Total	719	76	795

Table 21 - Product rule works nearly flawlessly (correlation of 0).

An obvious conclusion here is that the Product Rule performance depends on the pair of characteristics chosen, ranging from near perfect results (the product rule holds) to results that would greatly mislead if the product rule were mechanically applied. A histogram of all the pairs of possible correlations for the database is given in Figure 2. From the histogram and frequency table, 97.01% of all the 94,830 combinations of features have a coefficient of correlation between -0.2 and +0.2. This leaves a very small percentage (2.99%) of combinations of features that should not be considered when applying the product rule, with most of these combinations arising from two features within the same letter. Limiting the correlation to one feature per character (thus excluding intra-character interdependence), 97.39% of the combinations have a coefficient correlation between -0.2 and +0.2 and 2.61% of the correlations coefficients are greater than 0.2 in absolute value. It should be emphasized that this analysis was limited to pairs only and not to larger sets of features.

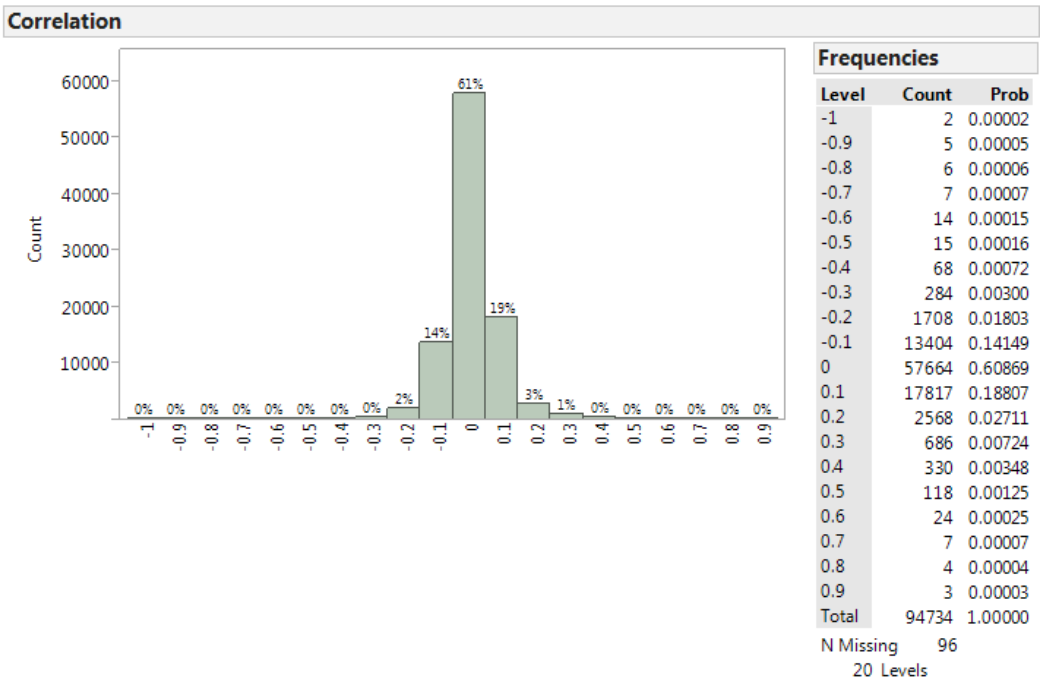

Figure 4 - Histogram of All Possible Correlations Among the 435 Features.

The database using the 880 cursive specimens can be used for numerous other investigations, such as examining the product rule with three or more characteristics. Although the 880 cursive specimens may seem to be a substantial reduction over the 1500 specimens collected (albeit not fully examined), this cleaned database enjoys a demographic basis that adequately matches the US resident population relative to our initial quotas. We have merely begun in terms of possible analyses but the database provided as a deliverable in this project is offered for others to assess for their particular issues. Additional specimens could certainly be reviewed, keeping in mind to continue improving the quotas.

Hand Printed

Since feature presence can be thought of as an indicator variable (i.e., either 1 for present/TRUE or 0 for absent/FALSE), we can further consider each variable as having a binomial distribution. The individual samples are effectively independent (no one was copying from another contributor) and we posit the overall probability of presence as an unknown parameter p_i, where i goes from 1 to 351 (the number of printed characteristics). The standard deviation of the proportion estimator is given as $[p_i (1- p_i)/839]^{0.5}$. (The denominator 839 will vary slightly depending on the actual count of the letter as referenced in the third column of Table 17.)This expression attains a maximum of 0.0173 for $p_i = 0.5$ and is rather less as the proportion approaches either 0 or 1.

Using standard statistical methodology the 95% confidence limits of the frequency occurrence proportions in Table 17 will be a range from -2 sig to +2 sig of those proportions

. "Sig" is short for "sigma" and is defined as meaning the estimated standard deviation. The term "95% confidence limits" is defined as meaning that there is a 95% confidence level that the frequency occurrence will fall within the calculated range (which is the standard for reaching statistically valid results for an entire population in a sampling environment). By applying the standard deviation estimator equation $[p_i (1- p_i)/839]^{0.5}$ to any of the hand printed frequency occurrence proportions one can easily calculate the 95% confidence limits by adding and subtracting the standard deviation times two. (The denominator 839 will vary slightly depending on the actual count of the letter as referenced in the third column of Table 17.) An example of this calculation is found in the cursive results section of this report. There have been previous claims that unless the entire population is tested one cannot have statistically valid calculations as to frequency occurrence. Those claims are incorrect and contrary to standard statistical practices.

As noted earlier in this report, Huber and Headrick (1999) describe *qualitatively* various demographic features that influence handwriting in general. With the study sample established and in conjunction with the associated demographics, we can *quantitatively* assess character features as a function of age, gender, ethnicity, education level, location of cursive training, and handedness. For each combination of character feature (351 for printed) and demographic (6 possibilities), the association as measured by Fisher's exact test has been run. A significant association implies that the demographic variable influences the presence/absence of a feature. Table 5 summarizes the results for all 351 features across the 6 demographic variables. The demographic row total basis values vary depending on the number of unknowns for the demographic category, a sub-category deliberately not included (e.g., Native American for Ethnic or ambidextrous for Right or Left which have small sample sizes), or some specimens not examined for a given letter and demographic (hence, the range of row totals provided).

p-value range	Age	Location 2nd or 3rd Grade	Gender	Education	Ethnic	Right or Left
<.0001	11.9%	0.3%	5.0%	1.4%	6.4%	0.6%
<.001	16.3%	0.8%	9.1%	4.2%	6.9%	1.1%
<.01	24.9%	1.9%	18.0%	13.0%	10.0%	1.4%
<.05	34.9%	3.9%	24.1%	21.1%	18.8%	5.0%
<.1	41.8%	11.6%	31.3%	27.1%	25.2%	9.4%
>.95	3.3%	27.4%	17.7%	2.2%	10.0%	30.7%
1	2.8%	27.4%	17.7%	2.2%	6.9%	30.7%
Demographic Row Total Range	606-810	501-689	610-815	602-805	587-784	607-810

Table 19 - Percentage of 351 Features Having Indicated p-value Range.

Table 19 is arranged from left to right according to the strength of association (stronger to weaker) that was observed with the cursive specimens. With regard to the printed specimen situation, the location of 2nd/3rd grade education is no longer associated with features. Overall, age has the greatest bearing on the number of features. Over a quarter of the features considered exhibit an effect on the presence/absence due to age of provider. Gender and education also exhibit a significant signal for many of the features (many more than would be expected due to chance alone). The ethnic category (restricted to Caucasian, African American, Hispanic and Asian) also influences a number of features presence/absence. Handedness continues to have little to do with influencing presence/absence with percentages matching those that would be observed due to chance alone.

Closeness of writers based on presence/absence of features. With the data base, we examined those specimens that had no missing characteristics (423 specimens). Using cluster analysis on this group of complete entries, it was found that the "closest" specimens in terms of presence/absence of features were original specimens (#361, #399) with 21 features that differed, followed by original specimens (#348, 1128) with 23 features that differed. Any other two specimens have more than 23 features that are different with typically 80 to 100 features not in common. It must be emphasized that even a complete match does not suggest that writers are one and the same or that these collection of presence/absence features are sufficient to completely determine authorship. Two specimen providers could conceivably exhibit the same set of database responses. However, upon examination by an expert forensic document examiner, additional subtleties and variances within the present/absent modes would be noted to distinguish authorship. In other words, reliance solely on the degree of matching is contrary to the intent of our research and to appropriate forensic document examination methodology (see Figure 6).

Product Rule Analyses. The product rule has been defined as "the probability of concurrence of all the events is equal to the continued product of the probabilities of all the separate events" (Osborn (1929) referencing Professor Simon Newcomb). A key factor in using the product rule is that each event must be independent of the other events in order to be applicable. It is recognized as a convenient tool if in fact it applies to this study. The data sets allow us to consider numerous instances to test the appropriateness of the product rule with respect to presence or absence of combinations of characteristics. Since there were 361 features for this analysis (10 features pared subsequent to interdependency testing), there were a total of 64,980 possible pairs of features that could be considered.

A histogram of all the pairs of possible correlations for the database is given in Figure 5. Overall, there are 98.55% of all the 64,980 combinations of features that have a coefficient of correlation between -0.2 and +0.2. This leaves a very small percentage (1.45%) of combinations of features that should not be considered when applying the product rule, with most of these combinations arising from two features within the same letter. Limiting the correlation to one feature per

character (thus excluding intra-character interdependence), 98.96% of the combinations have a coefficient correlation between -0.2 and +0.2 and 1.04% of the correlations coefficients are greater than 0.2 in absolute value. It should be emphasized that this analysis was limited to pairs only and not to larger sets of features.

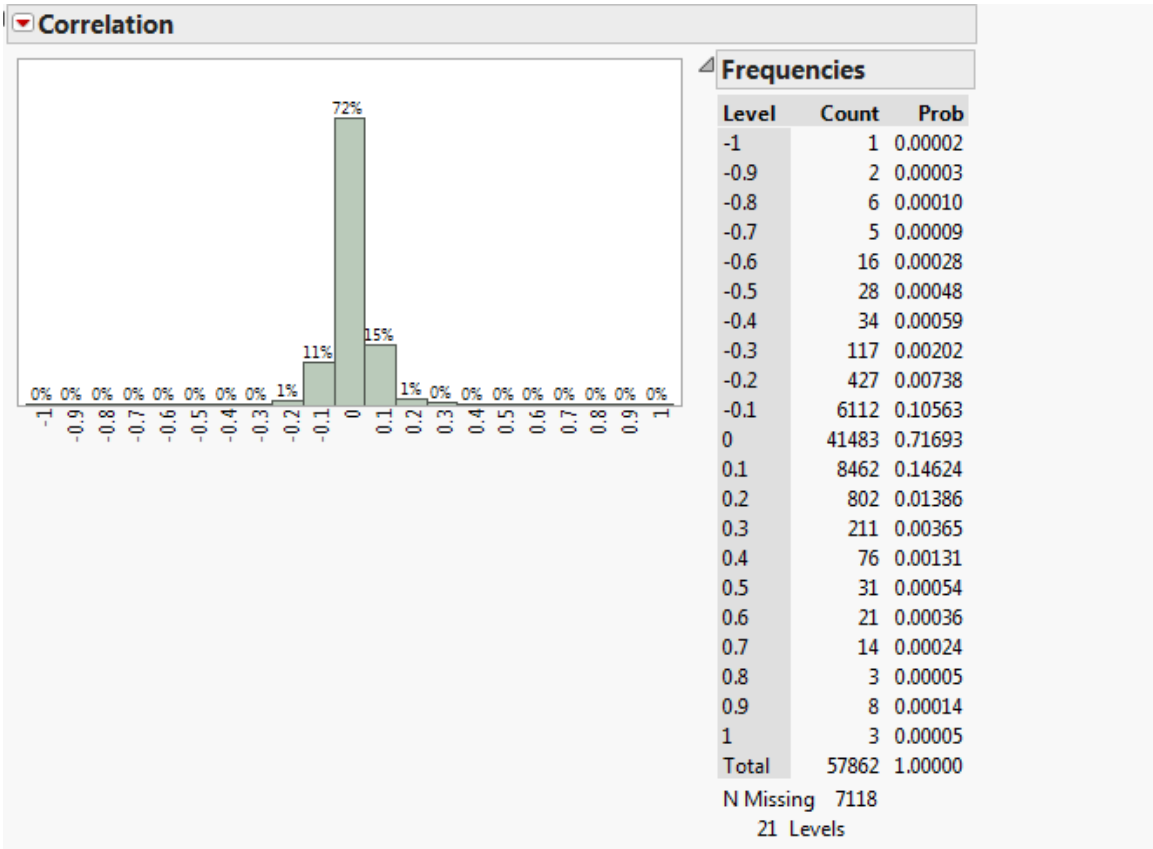

Figure 5 - Histogram of All Possible Correlations among the 351 hand printed features, 19 of which show no variability, explaining the missing 7,118 coefficients of correlation.

The data base using the 839 printed project specimens can be used for numerous other investigations, such as examining the product rule with three or more characteristics. Although the 839 printed project specimens may seem to be a substantial reduction over the 1500 specimens collected (albeit not fully examined), this cleaned data base enjoys a demographic basis that adequately matches the US resident population relative to our initial quotas. We have merely begun in terms of possible analyses but the data base provided as a deliverable in this project is offered for others to assess for their particular issues. Additional specimens could certainly be reviewed, keeping in mind to continue improving the quotas.

5. Conclusions

The original goals of this project were;

1. Develop statistically valid frequency occurrence proportions of handwriting and hand printing characteristics based on specimen samples from throughout the United States;
2. Provide practitioners of forensic document examination with statistical basis for reliability and measurement validity to accurately state their conclusions and assess complexity; and
3. Provide courts with the reliable data needed to understand the underlying statistical basis for the conclusions.

Data collection is the end product for the project. The frequency occurrence proportions have been developed and are listed in Tables 16 and 17. Forensic document examiners now have hundreds of statistically valid proportions available for their use.

Forensic document examiners have many options for use of this data in their daily practice. The most obvious application of this data is to be able to provide a baseline figure of distinctiveness for any cursive or hand printed entry. This data could also be useful for objectively assessing complexity of what appears to be a generic style of writing. One could also apply the proportions as objective guidance to the overall complexity of a small amount of writing. This information might also prove useful as a statistically-based element for estimate of confidence regarding conclusions. The entire profession must now begin a dialogue for the use of this material and its interjection into the profession methodologies and assistance tools.

Courts have been requesting statistical underpinnings for the basis of expert conclusions for years. Forensic document examiners now have that information available to share in the courtroom and assist in educating the judiciary with scientifically objective data that provides an appreciation of the statistical heterogeneity in any given handwritten entry that may be a central issue in either civil or criminal litigation.

Statistical studies in this report have concluded as to the very high degree of independence of cursive and hand printed entries and the basis for use of the product rule in overall writing probabilistic individuality assessment.

Forensic document examiners have had detailed qualitative information concerning extrinsic and intrinsic effects on handwriting. Now the profession has quantitative data concerning a few of the more common factors that affect handwriting.

Not one set of samples in either cursive or hand printed specimens contained the exact same present/absent results. This distinguishing factor is solely based on

presence/absence of characteristics and does not even take into account overall variations and more subtle variations within common overall designs.

The authors of this report are currently working on the production of a canned query containing the proportion results listed in Tables 16 and 17 in the format of the original database which contain both a checkbox system of presence/absence and illustrative images to accompany the feature descriptions (see Figure 1). The purpose of this query will be to allow forensic document examiners to input the presence of features germane to their examination and receive a boilerplate report listing the feature description and the individual frequency occurrence proportion. In addition the report will contain the result of the application of the product rule to the data. Until such time, a downloadable version of the database will be available at the website for the National Center for Forensic Science at ncfs.ucf.edu. The database is too large to print and incorporate into this report. However, it will be helpful in fully understanding the scope of each feature description and its download is recommended to fully appreciate the content of Tables 16 and 17.

There is a high potential for misuse of the information in this project. As such many cautionary comments are warranted.

It should be understood that the scope of characteristics examined by forensic document examiners in the course of any examination will far exceed the numbers presented in this project by many factors. This project has just scratched the surface of the detail that is reviewed and is designed to provide an appreciation of the probabilistic level of individuality in handwriting. Forensic document examiners should not be limited solely to the features listed in this project as doing so would be a specific misuse of the intent of this project and the scope of standard document examinations. Huber and Headrick (1999) list the following twenty-one discriminating elements scrutinized in a full and complete forensic document examination:

1. Arrangement
2. Class of Allograph
3. Connections
4. Design of Allographs and their Construction
5. Dimensions
6. Slant or Slope
7. Spacings
8. Abbreviations
9. Alignment
10. Commencements and Terminations
11. Diacritics and Punctuation
12. Embellishments
13. Legibility or Writing Quality (Including Letter Shapes)
14. Line Continuity
15. Line Quality

16. Pen Control
17. Writing Movement
18. Consistency or Natural Variation
19. Persistency
20. Lateral Expansion
21. Word Proportions

This project in no way promotes or describes methodology for the comparative examination of handwriting based solely on the results of this research. Should an individual attempt to simulate or trace another's writing, it would be expected to find a significant degree of agreement in the presence/absence of the features described in this report. However, the vital features of line quality, blunt ending strokes, hesitations, pen lifts, and other features of simulations or tracings (that are used by qualified forensic document examiners to expose such activity) are not an aspect of this study

One should not apply any inverse application to the frequency occurrence proportions. If the presence of a characteristic has a frequency occurrence proportion of 0.25, it cannot be assumed that the absence of this characteristic has a frequency occurrence of 0.75. The reason for this is variation in handwriting. This study not only applies a present/not present format for establishment of characteristic frequency but also applies presence priority. Per the example if the character being reviewed was present once but absent one hundred times within the handwriting specimen, the database box would reflect the presence of the characteristic. Likewise if the feature at issue was whether a specific characteristic was not present then one could not apply the inverse of the frequency occurrence proportion if the feature was present for the same reason.

This project provides no data on which to advocate computerized handwriting comparisons. This project recognizes that there are computer-based tools that can assist a properly trained and fully qualified forensic document examiner but only as a tool and not as competition or as a replacement.

There are not necessarily homogenous reasons for the notation that a characteristic is present or absent as illustrated in Figure 6. As such, the presence or absence of any characteristic just begins to illustrate the level of differences in handwriting characteristics and provides an appreciation for the level of distinction to any given characteristic. The reader should understand that there are other factors that provide additional contributions to the determined level of heterogeneity of handwriting based, for instance, on the different reasons for which a box was checked or not checked in the database

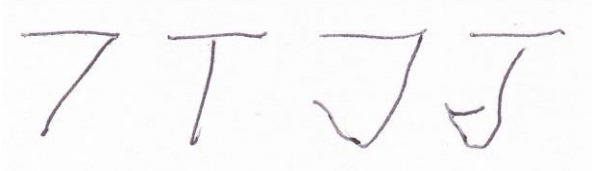

Figure 6 – Cursive Upper Case T (CUCT) Feature 14 "Cap is approximately straight" applies to each of the above versions of a cursive upper case "T". This figure illustrates the non-homogenous reasons that boxes are checked and why presence/absence is a small aspect of individuality and comparison assessment by forensic document examiners.

6. Future Research Directions

There are a number of obvious next steps in progressing this research and its impact on forensic document examination. The most direct impact would be to increase the population sample and increase the features in the database. Doing so would result in better and more comprehensive data.

The forensic science community must now weigh the significance and use of the data in daily examination work. While this project can provide guidance and make recommendations it is only after profession-wide dialogue and testing is initiated will practical applications become accepted methodologies.

This project's influence is limited to the United States as the population sampling is based on the U.S. population and the Latin alphabet. Foreign countries will need to apply their own set of population sampling strata then select and test location-specific handwriting specimens in keeping with the population in order to have data relevant to their location. It is hoped that this project can provide the model from which many countries will mirror this project.

Other analyses could be considered beyond those conducted for this project. Presently, there is considerable interest in developing likelihood ratios for adjusting the posterior distributions in legal proceedings. Adapting the work of Davis et al. (2012) to take advantage of the database is a worthwhile extension

7. References

ASTM-E105-04 (2004). Standard Practice for Probability Sampling of Materials. ASTM International, West Conshohocken, PA.

ASTM-E141-91 (1991). Standard Practice for Acceptance of Evidence on the Results of Probability Sampling. ASTM International, West Conshohocken, PA.

ASTM-E1658 (2008). Standard Terminology for Expressing Conclusions of Forensic Document Examiners. ASTM International, West Conshohocken, PA.

ASTM-E2290 (2007). Standard Guide for Examination of Handwritten Items. ASTM International, West Conshohocken, PA.

Beacom M (1960). "A Study of Handwriting by Twins and Other Persons of Multiple Births", *Journal of Forensic Sciences,* **5**(1).

Bishop B. (2012). "Frequency of Selected Hand Printing Characteristics Occurring within a National Population", *Journal of American Society of Questioned Document Examiners,* **15**(2).

Boot D. (1998). "Degree of Similarity in the Handwriting of Twins", *Journal of American Society of Questioned Document Examiners,* **1**(2).

Boulanger-Carey M., Chen, H.T., Descloux, A., Ingle, J.F. and Park, K.I. (1984). "1982/83 End Office Connections Study: Analog Voice and Voiceband Data Transmission Performance Characterization of the Public Switched Network," *AT&T Bell Laboratories Technical Journal*, **63**(9): 2059-2119.

Boulanger, M., Johnson, M.E., and Vastrick, T. (2013). "Development of an Extensive Forensic Handwriting Database—Statistical Components," Proceedings from the Seventh Workshop on Simulation, Rimini, Italy, May 21-25, 2013.

Bundeskriminalamt – Forensic Information System for Handwriting.

Chamberland G. and Ghirotto M. (1990). "A Statistical Approach to Handwriting Comparison: Search for Characteristics that are Usable in the General Lineal Model", Canadian Society of Forensic Sciences Presentation.

Cusak C and Hargett J (1988). "A Comparison Study of the Handwriting of Adolescents" American Academy of Forensic Sciences Presentation.

Davis LJ, Saunders CP, Hepler A, Buscaglia J. (2012). "Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios", *Forensic Science International*, **216**:146-57.

Durina M and Caligiuri M (2009). "The Determination of Authorship from a Homogenous Group of Writers", *Journal of American Society of Questioned Document Examiners,* **12**(2).

Dyer A, Found B, Rogers D. (2006). "Visual Attention and Expertise for Forensic Signature Analysis, *Journal of Forensic Sciences,* **51**(6).

Eldridge M, Nimmo-Smith I, Wing A, and Totty R (1984). "The Variability of Selected Features in Cursive Handwriting: Categorical Measures," *Journal of Forensic Science Society,* **24:** 179-219.

Fienberg, S. (2007). "Editorial: Statistics and Forensic Science," *The Annals of Applied Statistics*, **1**: 285-286.

Finneran, K. (2003). "Flaws in Forensic Science," *Issues in Science and Technology*, **20**, Fall Issue.

Gamble D. (1980). "The Handwriting of Identical Twins," *Journal of Canadian Society of Forensic Sciences,* **13**(2): 11-30.

Hilton, O. (1958). "The Relationship of Mathematical Probability to the Handwriting Identification Problem" RCMP Seminar Presentation

Horton R. (1996). "A Study of the Occurrence of Certain Handwriting Characteristics in a Random Population," *International Journal of Forensic Document Examiners,* **2**(2).

Huber, R. and Headrick, A. (1999). *Handwriting Identification: Facts and Fundamentals*. CRC Press, Boca Raton, FL.

Huber R. (2000). "The Heterogeneity of Handwriting," *Journal of American Society of Questioned Document Examiners,* **3**(1).

ISO/TR 14468:2010(E), Selected Illustrations of Attribute Agreement Analysis, Geneva: International Standards Organization.

Kam, M., Wetstein J., Conn R. (1994). "Proficiency of Professional Document Examiners in Writer Identification," *Journal of Forensic Sciences,* 39(1): 5-14.

Kam, M., Fielding, G., Conn, R. (1997). "Writer identification by professional document
Examiners," *Journal of Forensic Sciences,* **42**: 778–786.

Kam M, Fielding, G., Conn R. (1998). "Effects of Monetary Incentives on Performance of Nonprofessionals in Document-Examination Proficiency Tests," *Journal of Forensic Sciences,* **43**(5): 1000 – 1004.

Kam M., Gummadidala K., Fielding, G., Conn R (2001). "Signature Authentication by Forensic Document Examiners, *"Journal of Forensic Sciences,* **46**(4): 884-888.

Kam M. (2010). "Proficiency Testing and Procedure Validation for Forensic Document Examination," Report to Technical Support Working Group.

Kelly J (2002). "Habits Observed in Naturally Written Numbers," *Journal of American Society of Questioned Document Examiners.*

Kennedy, D. (2003). "Forensic science: Oxymoron?", *Science*, **302**: 1625.

Lee C and Abbey R (1922). *Classification and Identification of Handwriting.* D Appleton and Co., New York, NY.

Livingston O. (1959). "A Handwriting and Pen-Printing Classification System for Identifying Law Violators", *Journal of Criminal Law, Criminology and Police Science,* **49.**

Livingston O (1963). "Frequency of Certain Characteristics in Handwriting, Pen Printing of Two Hundred People", *Journal of Forensic Sciences,* April 1963.

Muehlberger, R., Newman, K (1977). "A Statistical Examination of Selected Handwriting Characteristics." *Journal of Forensic Sciences,* **22** (1): 206-215

Nicholson P (1999). "A System for the Classification of Block Capital Letters", *International Journal of Forensic Document Examiners,* **5:** 138-145.

NRC (2009). *Strengthening Forensic Science in the United States: A Path Forward*. National Research Council, Washington, DC, Committee on Identifying the Needs of the Forensic Sciences Community and Committee on Applied and Theoretical Statistics.

Osborn, A. (1929). *Questioned Documents, 2d Ed.,* Patterson Smith, Montclair, New Jersey

Popkiss A, Moore J (1945). "Handwriting Classification", *Police Journal*, **18**: 39-55.

Savoie K (2011). "The Frequency of Occurrence of Specific Handwriting Characteristics within a Limited Population", *Journal of American Society of Questioned Document Examiners,* **14**(2).

Schuetzner, E. (1999). "Class Characteristics of Hand Printing", *Journal of the American Society of Questioned Document Examiners,* **2** (1)

Schuetzner, E. (2000). "Class Characteristics of Numbers", American Academy of Forensic Sciences Presentation

Shiver F (1996). "Case Report: The Individuality of Handwriting Demonstrated Through the Field Screening of 1000 Writers," American Society of Questioned Document Examiners Presentation.

Sita J., Found, B., Rogers, D. (2002). "Forensic handwriting examiner's expertise for signature comparison", *Journal of Forensic Sciences,* **47**: 1117–1124.

Srihari, S.N., Leedham, G. (2003). "A survey of computer methods in forensic document examination", *Proceedings of the 11th International Graphonomics Society Conference*, Scottsdale, AZ: 278–281.

Srihari S.N., Cha, S.H., Arora, H., Lee, S. (2002). "Individuality of Handwriting", *Journal of Forensic Sciences*, **47**: 1-17.

Srihari S.N., Huang C, Srinivasan H (2006). "On the Discriminability of the Handwriting of Twins," *Journal of Forensic Sciences* **53**(2).

Srihari S.N. (2010). "Computational Methods for Handwritten Questioned Document Examination," NIJ Report 232745.

Srihari S.N. (2013). "Statistical Examination of Handwriting Characteristics Using Automated Tools," NIJ Report 241743.

SWGDOC (2012). Standard Terminology for Expressing Conclusions of Forensic Document Examiners. Scientific Working Group – Forensic Document Examination.

SWGDOC (2012). Standard for Examination of Handwritten Items. Scientific Working Group – Forensic Document Examination.

United States Secret Service – Forensic Information System for Handwriting.

Vastrick, T. (1998). "The Uniqueness of Handwriting", *Journal of American Society of Questioned Document Examiners*, **1**(1): 4-7.

Welch J. (1996). "A review of Handwriting Search Cases as an Indicator of the Individuality of Handwriting", American Society of Questioned Document Examiners Presentation.

Acknowledgements

The authors wish to acknowledge the assistance of a host of individuals that contributed to many phases of this project.

The administrative offices of the National Center for Forensic Science (NCFS) on the University of Central Florida campus in Orlando, FL.

Dr. Michael Sigman, PhD, NCFS Director**; Dr. Jack Ballantyne, PhD,** Former Interim Director NCFS**; Carrie Whitcomb, MS,** Former Director NCFS - for their tenures as Primary Investigator of this project.

Ms. Judith Stout, Assistant Director Research Programs and Services NCFS – all things administrative (that one must-have person with all the answers)

Kathleen Storer and Karen Runyon (posthumous), Forensic Document Examiners – with Schuetzner and Vastrick comprised the core of the genesis of this project.

Heather Burske, Database Specialist – developed a user-friendly database specifically for this project.

Dennis Mooney, Miriam Angel, Patricia Zippo, Janis Tweedy, Howard Seiden, Diane Flores, Raesin Caine and Larry Olson, Forensic Document Examiners - assisted in the classification or collection processes.

Kevin Kulbacki, Forensic Document Examiner **–** began as student intern, also served as formal presenter of updates at forensic conferences.

Inalvis Alvarez-Fernand, Tiffany Benton, Christopher Cropanese, Ryan Dungan, David Evander, Michael Hathcock, Hannah Hines, Shelby Khandasammy, Jessica Kindell, Lauren McCool, Colleen McGuire, Brittany Motta, Ali Norero, Nicole Pike, Flor Rodriguez, Stephanie Rosser, Martha Sillars, La'Quida Smith, Jessica Sprague, Luke Spratt, Melissa Sprung, Zoraida Torrelli, Fiorella Travi, Annette Way Student Interns - instrumental in the successful completion of this project in many different ways.

Gerald Mattson, PhD and **Mattheu Miller, MS** – Forensic Science Curriculum, Department of Chemistry, University of Central Florida - assistance in recommending student interns.

Appendices

Appendix 1 - Specimen Letter[1]

From: Jim Elder
829 Loop Street, Apt. 300
Allentown, New York 14707

To: Dr. Bob Raj Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started about six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event. Kate was overworked. But she enjoyed her job and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!

Jim

[1] Slightly Modified from Srihari et al., 2002.

APPENDIX 2 – Assistant Illustration Form Samples

From: Jim Elder
829 Loop Street, Apt. 300
Allentown, New York 14707

b

To: Dr. Bob Raj Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started about six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event. Kate was overworked. But she enjoyed her job and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

From: Jim Elder
829 Loop Street, Apt. 300
Allentown, New York 14707

w

To: Dr. Bob Raj Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started about six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event. Kate was overworked. But she enjoyed her job and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

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

Appendix 3 - Data Preparation Process and Results - Cursive

As is customary with real data applications, considerable data preparation is required prior to conducting any formal quantitative analyses. Our analyses proceed based on a cleaned version of the full data set. Wherever possible, we could return to the original specimens for re-examination and clarification, but in lieu of that very time-consuming and not necessarily fruitful route we make tactical adjustments as explicitly indicated in this Appendix. The raw data was provided for each demographic variable in turn and then indicate the categories to which these are mapped. The goal is to preserve the assessments of as many specimens, as possible with transparency on the mapping decisions.

For cursive specimens and the variable age, there were 18 entries requiring some attention. (Given the level of effort in examining the documents, including the results of the examination with possibly unknown for a specific demographic response is preferred.) The specific responses for age are illustrated in Table A1.

Count	Reported	Coded		Count	Reported	Coded
3		U		26	54	60
1	`36	36		30	55	60
4	0	U		29	56	60
1	1	U		17	57	60
1	100	60		21	58	60
1	16	24		22	59	60
1	17	24		22	60	60
84	18	24		16	61	60
86	19	24		18	62	60
68	20	24		24	63	60
73	21	24		19	64	60
50	22	24		19	65	60
34	23	24		11	66	60
35	24	24		8	67	60
41	25	24		15	68	60
27	26	24		14	69	60
31	27	24		8	70	60
31	28	24		7	71	60
20	29	24		5	72	60
26	30	24		4	73	60
28	31	40		9	74	60
22	32	40		4	75	60
18	33	40		4	76	60
24	34	40		12	77	60
20	35	40		8	78	60
21	36	40		10	79	60
17	37	40		1	7923	60
25	38	40		6	80	60
19	39	40		3	81	60
24	40	40		1	82	60
19	41	40		4	83	60
21	42	40		3	85	60
20	43	40		2	86	60
13	44	40		2	87	60

17	45	40		2	88	60
27	46	40		1	89	60
20	47	40		1	FALSE	U
27	48	40		1	female	U
18	49	40		3	n/a	U
28	50	40		1	N/A	U
18	51	60		1	Not given	U
13	52	60		1	not specified	U
25	53	60				

Table A1 – Ages Reported Raw Data

Following the recoding to meaningful categories to match the definitions of Huber and Headrick, aggregation occurs as illustrated in Table A2.

Age Re-coded	count	%	Quota %
24	608	40.5%	> 20%
40	429	28.6%	> 30%
60	464	30.9%	> 30%
Unknown	16		

Table A2 – Ages Reported and the Recoded Values

For gender, Table A3 summarizes the raw entries.

Count	Recorded	Re-coded
1		U
1	0	U
1	female	F
975	Female	F
1	male	M
532	Male	M
2	n/a	U
1	N/A	U
1	NA	U
1	not specified	U
1	TRUE	U

Table A3 - Gender Reported Raw Data

The corresponding aggregated results for gender are illustrated in Table A4.

Sex recoded	count	%	Quota %
F	976	64.7%	> 40%
M	533	35.3%	> 40%
U	8		

Table A4 - Gender Reported and the Recoded Values

Self-identified ethnic classifications were recorded in various ways, as reported in Table A5.

Count	Old Value	New Value
12		U
1	0	U
7	African-american	African-American
95	African-American	African-American
1	asian	Asian
51	Asian	Asian
2	black	African-American
1	Black/Native American/Caucasian	Mixed
1187	Caucasian	Caucasian
1	caucasian/african american	Mixed
1	caucasian/hispanic	Mixed
1	caucausian	Caucasian
1	FALSE	U
7	hispanic	Hispanic
100	Hispanic	Hispanic
1	hispanic/asian	Mixed
1	male	U
1	middle eastern	Other
5	Middle Eastern	Other
2	mixed	Mixed
1	multiracial	Mixed
2	n/a	U
2	N/A	U
1	NA	U
3	Native American	Native American
1	Not Filled In	U
1	Not given	U
1	not specified	U
3	other	U
7	South Pacific	South Pacific
2	unknown	U
15	white	Caucasian

Table A5 - Ethnicity Reported Raw Data

Aggregations and re-classifications were performed. The results are illustrated in Table A6.

Ethnic	count	%	Quota %
African-American	104	7.0%	> 10%
Asian	52	3.5%	> 4%
Caucasian	1203	80.8%	> 55%
Hispanic	107	7.2%	>11%
Mixed	7	0.5%	
Native American	3	0.2%	
Other	6	0.4%	
South Pacific	7	0.5%	
U	28		

Table A6 - Ethnicity Reported and the Recoded Values

Right and left handedness did not escape the vagaries of data entry issues, as seen in Table A7.

Count	Old Value	New Value
1		U
1	]=\ffffffffffffffffffffffffffffffff	U
1	0	U
1	ab	U
4	Ambedextrous	R+L
1	left	L
131	Left	L
2	n/a	U
1	N/A	U
1	NA	U
1	not specified	U
2	right	R
1369	Right	R
1	TRUE	U

Table A7 - Handedness Reported Raw Data

The recoded version of the handedness data set is illustrated in Table A8.

RorL RC	count	%	Quota %
L	132	8.8%	> 5%
R	1371	91.2%	> 75%
R+L	4		
U	10		

Table A8 – Handedness Reported and the Recoded Values

For education, the raw data and the recoded values are given in Table A.5.

Count	Old Value	New Value
15		U
1	0	U
1	14	U
1	14 years	U
1	16	U
1076	Above High School Diploma	HS plus
36	Advanced Degree	HS plus
1	Assiociate's Degree	HS plus
2	Associate's Degree	HS plus
4	Bachelor of Science	HS plus
58	Bachelor's Degree	HS plus
1	College	HS plus
1	college graduate	HS plus
1	College Graduate	HS plus
1	College Graduate / BS	HS plus
85	Diploma	HS
1	Doctorate	HS plus
1	Florida	U
1	Graduate Degree	HS plus
212	High School Diploma or Less	HS
1	IN	U
1	K thru 12 + 1 semester of college	HS plus
1	MACJ	U
5	Master's Degree	HS plus
1	Masters Degree	HS plus
1	More than High School	HS plus
2	n/a	U
1	N/A	U
1	NA	U
1	not specified	U
1	Some college	U
1	TRUE	U

Table A9 – Level of Education Reported Raw Data

The recoded version for education is illustrated in Table A-10.

Educ RC	count	%	Quota %
HS	297	20.0%	>30%
HS plus	1191	80.0%	>50%
U	29		

Table A10 – Level of Education Reported and Recoded Values.

The location of second/third grade (where and when cursive would normally have been taught) generated a very large number of responses. Following the identification of obvious US locations or unknown responses (handled similarly as in the other cases presented previously in this Appendix), the disposition of responses is illustrated in Table A11. Samples collected listed 47 of the 50 states as the location of second/third grade education.

Location	#		Location	#
AF	1		Lima, Peru	1
Africa	2		Mexico	2
American Samoa	1		Monterrey, Mexico	1
Bayamon, PR	1		Montreal, Canada	1
Belarus	1		Muscat, Oman, ABA	1
Canada	2		New Zealand	3
Caymans	1		Newfoundland	1
Chile	2		Nicaragua	1
China	1		Nicaragua/Costa Rica	1
CN	2		Odessa, Ukraine	1
Colombia	2		Peru	1
Cuba	4		Philippines	2
Cyprus	2		Quebec	1
Dominican Republic	2		Russia	1
Edmonton, Alberta	1		Saskatchewan	1
England	1		Slovakia	1
France	2		Spain	1
Freeport, Bahamas	1		Taiwan	1
Germany	2		The Netherlands	1
GU	1		Toronto, Canada	1
Guam	1		UK	6
Haiti	1		Ukraine	1
Honduras	1		Unk	202
Hong Kong	1		US	1236
India	1		Vietnam	2
Iran	1		WC	1
Jamaica	2		Xiahong	1
Japan	1		Yakota Afb, Japan	1
Kenya, Africa	1		Yenn	1

Table A11 - Location of Educational Training at the 2nd / 3rd Grade Level Raw Data

Summarizing into the primary categories yields the results illustrated in Table A12.

2nd/3rd Grade	Count		

Education Location		%	Quota %
US	1236	94.0%	> 70%
Not US	79	6.0%	> 10%
Unk	202		

Table A12 – Location of Educational Training at the 2nd / 3rd Grade Level Reported and Recoded Values.

The results in this Appendix guided the selection of the project sample used for analyses as described in Table 16 of this report.

Appendix 4 - Data Preparation Process and Results – Hand Printed

The assessments of the printing specimens required data preparation work analogous to that given for the cursive writing specimens. Although the cleanup operations and re-coding of collected data is very similar to the printing case, the decisions made are provided here for transparency.

The variable age was coded into three age bins: 18-30 (coded as 24), 31-50 (coded as 40) and 51 and over (coded as 60) with the indeterminate entries coded as "U" as illustrated in Table A12.

Count	Reported	Coded		Count	Reported	Coded
3		U		26	53	60
1	`36	40		26	54	60
4	0	U		30	55	60
1	1	U		30	56	60
1	100	60		17	57	60
1	16	24		20	58	60
1	17	24		22	59	60
84	18	24		22	60	60
85	19	24		16	61	60
67	20	24		18	62	60
72	21	24		24	63	60
48	22	24		19	64	60
33	23	24		19	65	60
37	24	24		11	66	60
41	25	24		8	67	60
27	26	24		15	68	60
31	27	24		14	69	60
31	28	24		8	70	60
19	29	24		7	71	60
26	30	24		5	72	60
28	31	40		4	73	60
22	32	40		9	74	60
18	33	40		4	75	60
24	34	40		4	76	60
21	35	40		12	77	60
21	36	40		8	78	60
17	37	40		10	79	60
26	38	40		1	7923	60
19	39	40		6	80	60
24	40	40		3	81	60
19	41	40		1	82	60
21	42	40		4	83	60
20	43	40		3	85	60
12	44	40		2	86	60
17	45	40		2	87	60
27	46	40		2	88	60
20	47	40		1	89	60
28	48	40		1	female	U

19	49	40		3	n/a	U
28	50	40		1	N/A	U
18	51	60		1	Not given	U
13	52	60		1	not specified	U

Table A12 - Ages Reported Raw Data

Following the recoding to meaningful categories to match the definitions of Huber and Headrick, the following aggregation occurs:

Age Re-coded	count	%	Quota %
24	608	40.5%	> 20%
40	429	28.6%	> 30%
60	464	30.9%	> 30%
Unknown	16		

Table A13 - Ages Reported and the Recoded Values

For gender and printing assessments, Table A14 summarizes the raw entries.

Count	Recorded	Re-coded
1		U
1	0	U
1	female	F
974	Female	F
1	male	M
532	Male	M
2	n/a	U
1	N/A	U
1	NA	U
1	not specified	U

Table A13 - Gender Reported Raw Data

The corresponding aggregated results are illustrated below in Table A15.

Sex recoded	count	%	Quota %
F	975	64.7%	> 40%
M	533	35.3%	> 40%
U	7		

Table A14 - Gender Reported and the Recoded Values

Self-identified ethnic classifications were recorded in various ways, as reported in Table A15.

Count	Old Value	New Value
12		U
1	0	U
7	African-american	African-American
96	African-American	African-American
1	asian	Asian
51	Asian	Asian
2	black	African-American
1	Black/Native American/Caucasian	Mixed
1189	Caucasian	Caucasian
1	caucasian/african american	Mixed
1	caucasian/hispanic	Mixed
1	caucausian	Caucasian
7	hispanic	Hispanic
100	Hispanic	Hispanic
1	hispanic/asian	Mixed
1	male	U
5	Middle Eastern	Other
2	mixed	Mixed
1	multiracial	Mixed
2	n/a	U
2	N/A	U
1	NA	U
3	Native American	Native American
1	Not Filled In	U
1	Not given	U
1	not specified	U
2	other	U
7	South Pacific	South Pacific
1	unknown	U
14	white	Caucasian

Table A15 - Ethnicity Reported and the Recoded Values

Aggregations and re-classifications were performed and yielded results as illustrated in Table A16.

Ethnic	count	%	Quota %
African-American	105	7.0%	> 10%

Asian	52	3.5%	> 4%
Caucasian	1204	80.8%	> 55%
Hispanic	107	7.2%	>11%
Mixed	7	0.5%	
Native American	3	0.2%	
Other	5	0.3%	
South Pacific	7	0.5%	
U	25		

Table A16 – Ethnicity Reported and Recoded Values

Handedness did not escape the vagaries of data entry issues, as seen in Table A17.

Count	Old Value	New Value
1		U
1	]=\ffffffffffffffffffffffffffffffff	U
1	0	U
1	ab	U
4	Ambedextrous	R+L
1	left	L
131	Left	L
2	n/a	U
1	N/A	U
1	NA	U
1	not specified	U
2	right	R
1368	Right	R

Table A18 - Handedness Recorded Raw Data

The cleaned version of the data set is:

RorL Recoded	count	%	Quota %
L	132	8.8%	> 5%
R	1370	91.2%	> 75%
R+L	4		
U	9		

Table A18 - Handedness Reported and Recoded Values.

For education, the raw data and the recoded values are given in Table A19 and Table A20

Count	Recorded	Recoded
15		U

1	0	U
1	14	U
1	14 years	U
1	16	U
1074	Above High School Diploma	HS PLUS
36	Advanced Degree	HS PLUS
1	Assiociate's Degree	HS PLUS
1	Associate Degree	HS PLUS
2	Associate's Degree	HS PLUS
4	Bachelor of Science	HS PLUS
58	Bachelor's Degree	HS PLUS
1	College	HS PLUS
1	college graduate	HS PLUS
1	College Graduate	HS PLUS
1	College Graduate / BS	HS PLUS
85	Diploma	HS
1	Doctorate	HS PLUS
1	Florida	U
1	Graduate Degree	HS PLUS
212	High School Diploma or Less	HS
1	IN	U
1	K thru 12 + 1 semester of college	HS PLUS
1	MACJ	U
5	Master's Degree	HS PLUS
1	Masters Degree	HS PLUS
1	More than High School	HS PLUS
2	n/a	U
1	N/A	U
1	NA	U
1	not specified	U
1	Some college	HS PLUS

Table A19 – Level of Education Reported Raw Data

Educ RC	count	%	Quota %
HS	297	20.0%	>30%
HS PLUS	1191	80.0%	>50%
U	27		

Table A20 - Level of Education Reported and Recoded Values

The location of second/third grade (where and when handwriting would normally have been taught) generated a very large number of responses. Following the identification of obvious US locations or unknown responses (handled similarly as in the other cases presented previously in this Appendix), the disposition of responses is illustrated in Table A21. Samples collected listed 47 of the 50 states as the location of second/third grade education.

Location	#		Location	#
AF	1		Monterrey, Mexico	1
Africa	2		Montreal, Canada	1
American Samoa	1		Muscat, Oman, ABA	1
Bayamon, PR	1		New Zealand	2
Belarus	1		Newfoundland	1
Canada	2		Nicaragua	1
Caymans	1		Nicaragua/Costa Rica	1
Chile	2		Odessa, Ukraine	1
China	1		Peru	1
Colombia	2		Philippines	2
Cuba	4		Quebec	1
Cyprus	2		Russia	1
Dominican Republic	1		Santo Domingo, Dominican Republic	1
Edmonton, Alberta	1		Saskatchewan	1
England	1		Slovakia	1
France	2		Spain	1
Freeport, Bahamas	1		Taiwan	1
Germany	2		The Netherlands	1
GU	1		Toronto, Canada	1
Guam	1		U.K.	1
Haiti	1		Unknown	200
Honduras	1		US	1240
Hong Kong	1		Ukraine	1
India	1		Ukraine	1
Iran	1		Vietnam	2
Jamaica	2		WC	1
Japan	1		Xiahong	1
Kenya, Africa	1		Yakota Afb, Japan	1
Lima, Peru	1		Yemen	1
Mexico	2			

Table A21 - Location of Educational Training at the 2nd / 3rd Grade Level Raw Data

Summarizing into the primary categories yields the results illustrated in Table A22.

2nd/3rd Grade Education Location	Count	%	Quota %
US	1240	94.3%	> 70%

Not US	75	5.7%	> 10%
Unk	200		

Table A22 – Location of Educational Training at the 2nd / 3rd Grade Level Reported and Recoded Values

The results in this Appendix guided the selection of the sample used for analyses as described in Table 17 of this report.