

## ***FLASH ID***

***Forensic Language-Independent  
Analysis System for Handwriting  
Identification***

### **Objective Scoring Feasibility**

Submitted to:

**The Federal Bureau of Investigation  
Laboratory Division**

By:

**Sciometrics LLC  
DBA The Gannon Technologies  
Group**

CAGE Code: 3E6G9  
DUNS: 130785830  
14150 Parkeast Cir.  
Suite 140  
Chantilly, Virginia 20151  
(703) 793-2282  
Web: [www.gannontech.com](http://www.gannontech.com)

**July 24,2020 (Revised)**

TO: 15F06719F0001822  
CDRL: 004  
Task Order issued against IDIQ contract number  
DJF-15-1200-V-0007715

## 0.0 Table of Contents

### List of Contents

0.0 Table of Contents .....	1
1.0: Background and Introduction .....	2
1.1: Purpose of this Report .....	3
1.1: Technical Approach.....	4
1.3: Overview of Research Methods .....	11
2.0: Summary of Findings .....	12
2.1: Basic Principles for Scoring Fingerprints and Handwriting.....	12
2.2: Analysis with HPSEVERTY Procedure (SAS Function) .....	15
2.2: Deriving the Objective Score .....	17
2.3: Deriving the Objective Score for Handwriting .....	21
2.4 Other Considerations for Applying Objective Scoring to Handwriting.....	21
3.0 APPENDIX: Examples of rarity results from Fingerprint Study .....	23

## **1.0: Background and Introduction**

The Forensic Language-independent Analysis System for Handwriting IDentification (FLASH ID) is a fully functional software application designed to compare an unknown writing sample against a database of known reference writing. For each query based on a questioned document, FLASH ID returns a ranked list of the known samples based on embedded graphical similarity to the unknown document. FLASH ID works by maintaining a database of information derived from reference handwriting and determines whether a new, unidentified writing specimen, such as a questioned document, has a high degree of similarity to any of the writings in the database. FLASH ID operates on a conventional personal computer platform—including laptops.

Questioned documents subjected to biometric analysis are scanned and passed to FLASH ID as image files. Once the image has been captured, FLASH ID distills the biometric content from the handwriting, compares this content to reference samples stored in a database, computes scores representing biometric similarity between the questioned document and known documents, and compiles the results in a ranked list of all writers from the database. The writer at the top of this list bears the strongest biometric similarity to the writer of the captured specimen.

Currently, FLASH ID uses a competitive scoring matrix to determine the most similar writers from a database to the test document. Determining a likely candidate from analyzing the scoring involves identifying a significant gap in scoring from one writer to the next, indicating strong separation between handwriting characteristics. However, in actual casework, there are many factors that can limit the gap between scoring, such as quantity of writing or size of the writing in the sample.

FLASH ID calculates its comparative score based on several measurements taken of graphemes (small graphical elements that may be individual letters, parts of letters, or groups of letters) within a writing sample and how those grapheme measurements compare to grapheme measurements in other known writing samples. For every comparison between the test document and each reference document, scores are produced based on grapheme similarity/dissimilarity and the samples are ranked in order from the highest score in the first

position, to the lowest (many times negative) score in the last position. In the analysis of the FLASH ID recognition results, the user must examine the scores for a significant difference between sequentially ranked samples. A significant gap in the scoring suggests a significant difference in at least one grapheme measurement and, therefore, the candidate sample(s) ranked on the top edge of the gap are potentially associated. There are two issues associated with this approach for interpretation of scores: the true writer may not always be the highest scoring writer and the true writer may not be in the database. The former is a function of the algorithm performance; the latter is true of any automated system that uses comparison to databases that are not exhaustive of the population.

FLASH ID will always be limited by the contents of the databases; this is not an issue that can be addressed in the development of the software, but rather a function of the completeness of the databases. Limitations in score interpretation are mitigated in Latin-based languages by having the user compare the top-ranking writing samples to determine if an association does exist, or if the writer is low scoring or not in the database. This mitigation is currently not possible for non-Latin based languages, because there are no examiners in the FBI Laboratory who can perform a similar comparison; therefore, creation of a scoring system that would provide scores related to a confidence level of the association would allow the examiner to provide some interpretation to the results obtained from FLASH ID in non-Latin based languages. Hopefully, this will minimize the possibility of forwarding high-scoring candidate(s) when an association does not exist, or the writer is not in the database.

## **1.1: Purpose of this Report**

This report discusses a proposed method of scoring—originally developed for latent fingerprint matching—that would provide objective support in discrete and continuous scripts such as Latin-based languages (upper and lower case) in which forensic examiners are conducting handwriting comparisons. With current challenges in the legal system, any objective support that can be provided with the examiner’s report and opinion will strengthen the scientific foundation of the handwriting comparison.

- The purpose of this report is to explore the objective scoring approach evaluated under an NIJ grant for latent fingerprint comparison (using LatentSleuth’s fingerprint matching technology) for feasibility of use within the forensic handwriting domain (using FLASH ID).
- Our initial findings demonstrate the proof of concept (PoC) that utilizing the same general scoring algorithm from latent fingerprint research yields similar statistical properties when applied to handwriting. The approach is feasible within processing power currently applied to FLASH ID using data from the FBI500 writer database.

## 1.1: Technical Approach

A true objective score incorporated within FLASH ID will allow the FBI forensic document examiners to provide interpretation to the FLASH ID results in a statistical manner similar to the method used for presenting DNA evidence. Additionally, the objective score will open an avenue for examining writership of non-English documents and those with non-Latin characters—the handwriting on these documents are often not examined forensically if the examiners are not proficient in the language of the document.

The proposed Objective Scoring Algorithm uses large populations of known, non-matching references to create a predictive model *specific to a given test document* to compute the rarity of a given 1-1 comparison. This process leverages similarity scores between extracted features in the test document and a given reference, and has been successfully applied in latent fingerprint examination. The following two points compare the approach applied to fingerprints and handwriting.

- **Fingerprints:** the test sample is a latent, the features are small sample of level 2 ridge detail, and the scores are based on geometric similarity of the features between a latent and a reference, given an accurate ‘best warp’ between their two spaces (Latent Sleuth).
- **Handwriting:** the test sample is a test document, the features are graphemes, and the scores are the rewards output from discriminant analysis between a test and reference document (FLASH ID).

The discussion on the ensuing pages draws from techniques developed to improve the scoring capabilities of Sciometrics' LatentSleuth product. These techniques were tested and refined under a grant from the NIJ (Funding Opportunity Number NIJ-2017-11080).

The technique is a two-step process. Step one involves the creation of an Initial Model of Spurious Similarity using a small set known non-mate writers. This set is referred to as the Base Writer Set and can be the same writers used as the Base Writer Set in the current FLASH ID scoring system. For this study, we used a 50-writer Base Writer Set. In this step, each grapheme in the Test Document is compared in pairwise fashion to the writers in the Base Writer Set (who compete with each other for similarity to each test grapheme).

The 'Initial Model' is actually a set of similarity data completely internal to the Base Set. A 'General Similarity Model' will be created using a new (random) set of reference writers. We need a 'Base Set' (or an 'Initial Model') in order to generate similarity (to the QD) data for these new reference writers. Through competition with the Base Set writers for similarity to the QD (test probe), similarity data for this new writer is imbedded into the Base Set only similarity model. This data comprises an 'Objective Measure' of similarity to the QD for that reference writer. The 'General Similarity Model' created from this process is then a model of similarity relative to the measured similarity of the Base Set to the QD. The General Similarity Model is used to predict Rarity for a new reference writer whose similarity (to the QD) data is computed exactly as it was for each model building reference writer. The Rarity Prediction from the General Model will be independent of the specific Base Set used for and Initial Model. The Rarity Prediction will be a Data Base Random Match Rarity relative to the data base from which the model building reference writers are selected.

The Model coming out of this process provides the following capabilities for the evaluation of Reference Set Writers:

- An objective measure of the degree of similarity that an arbitrary writer exhibits to the Test Document;

# SCIOMETRICS

- The computational ability to compute the objective similarity measure for all writers in a very large set of known non-mates to create a basis for providing an assessment of the rarity of the objectively measured similarity of any Reference Writer.

Step two involves competing the Reference Set Writers against the Base Set Writers.

- (a) Each reference writer's similarity data is independently 'created' via competition with Base Set writers.
- (b) A General Model is created by doing this for a randomly selecting reference writers from a deliberate data base.
- (c) The General Model is the vehicle for predicting 'data base rarity' for a writer of interest whose similarity to the questioned document is computed exactly as it was for each reference writer used to generate the General Model.

We introduce the symbol omega,  $\omega$ , to represent a grapheme in the Test Document. For all pairs of Base Set Writers for any  $\omega$  in the Test Document, we compute Pairwise Competitions between the Base Set Writers for similarity to the test grapheme. We also compute Pairwise Competitions between each Reference Writer and all Base Set Writers for each grapheme,  $\omega$ , in the Test Document. Reference Set Writers do not compete against each other. Figures 1 and 2 display the pairwise competition patterns and formulas. Figure 1 describes pairwise competitions between Base Set Writers. Figure 2 describes pairwise competitions between Reference Set Writers and Base Set Writers. Each cell in the competitive matrix represents the Reward Score awarded to the writer based on the similarity of that writer's canonical variable for the grapheme class of the grapheme  $\omega$ . A canonical variable is a variable constructed from several other variables. In the case of handwriting, a canonical variable is created from geometric data such as physical angular measurements.







# SCIOMETRICS

For each Writer  $i$ , define  $S_i(\omega)$  as  $\log((R_i(\omega) + .5)/(C_i(\omega) + .5))$ . Heuristically,  $S_i(\omega)$  is the ‘Log Odds of Writer  $i$  Beating the Base Set.’

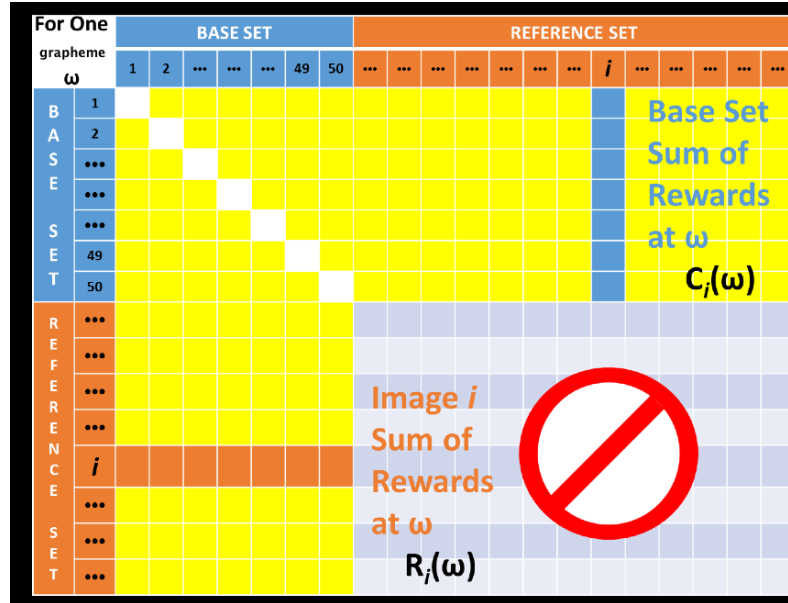


Figure 4: Writer  $i$  competing against the Base Set as a whole. In this example, the Writer  $i$  is a Reference Set Writer.

In order to achieve an informative statistic for the similarity of a writer to the Test Document, we perform the additional step of comparing  $S_i(\omega)$  and  $S_j(\omega)$  for all pairs of Writers  $i$  and  $j$ , where  $i$  and  $j$  cannot both be from the Reference Set of Writers (i.e., at least one of the  $i$  or  $j$  must be from the Base Set). Define

$$T_{i,j}(\omega) = \log\left[\frac{((R_i(\omega) + .5)/(C_i(\omega) + .5))}{((R_j(\omega) + .5)/(C_j(\omega) + .5))}\right]$$

Heuristically,  $T_{i,j}(\omega)$  is the logarithm of an odds ratio at the grapheme  $\omega$  level.

The  $S_i(\omega)$  scores measure how Writer  $i$  competes on similarity to the Test Document against the Base Set of Writers at the Test Document grapheme  $\omega$ ; and  $T_{i,j}(\omega)$  compares how Writers  $i$  and  $j$  compete against the Base Set of Writers at  $\omega$ .

The set  $\{T_{i,j}(\omega)\}_{i,j,\omega}$  are the results of competitions between all Writers  $i$  and  $j$  at graphemes  $\omega$  of the Test Document.

Using the  $\{T_{i,j}(\omega)\}_{i,j,\omega}$  data, we will explore methods for quantifying the overall performance of the Writers in the competitions and for ranking Writers according to their performance.

Our ultimate goal is to apply these methods to a large randomly selected set of known non-mate Writers to build a model that predicts the chance that the observed similarity to a Test Document by a new candidate Writer could be achieved by a random Writer. Base and Reference Set Writer comparisons are illustrated in Figure 5.

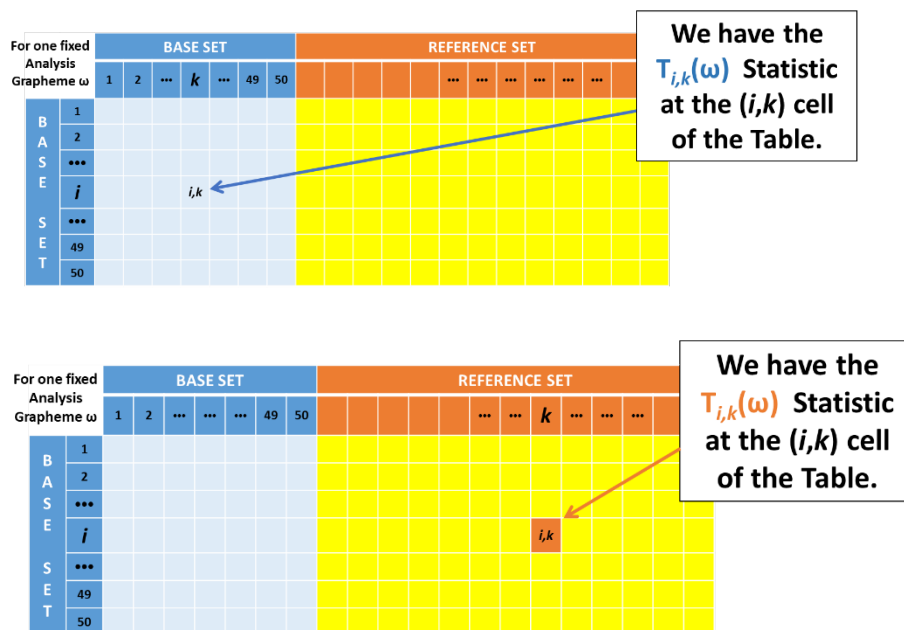


Figure 5: Writer Reference Set Competitions.

Positive  $T_{i,j}(\omega)$  Scores are called ‘Rewards.’ With Writers  $i$  and  $j$  from the Base Set, the Rewards for ‘Writer  $i$  versus Writer  $j$ ’ and for ‘Writer  $k$  versus Writer  $j$ ’ are compared, resolved and collected via the data analysis algorithms developed for the latent fingerprint case. These complex algorithms ultimately compute from Rewards, for each reference Writer  $k$ , a set of scores:  $z^j_{i,k}$ . When there are 50 Writers in the Base Set, then there are 2,450  $z^j_{i,k}$  scores for each reference Writer  $k$ . Each  $z^j_{i,k}$  score is an overall assessment of the similarity of Writer  $k$  to the test document. Collectively, the  $z^j_{i,k}$  scores comprise an objective measure of the similarity of Writer  $k$  to the test document. Further, when the 2,450  $z^j_{i,k}$  scores are *hierarchically* reduced to a single statistic, that statistic is an Objective Measure of the similarity of Writer  $k$  to the test document.

We demonstrate below the computation of a model of random similarity to a specific test document by computing the Objective Similarity Measure for many randomly selected non-mate writers.

## 1.3: Overview of Research Methods

As previously noted, the basis of the technique described here is derived from research performed on creating a statistical error prediction for an objective measure of similarity to an image of a latent fingerprint. There are a number of issues to be resolved in determining the effectiveness of this technique when applied to handwriting, including:

1. What is the proper number of Writers for the Base Set?
2. Should the Base Set be fixed for a given language or should it vary based on the Test Document?
3. What is the minimum number of known non-mate Writers required to build a model?
4. What is the minimum size (in graphemes) of a Test Document?

To transfer knowledge gained from fingerprints to handwriting, our research took the following steps:

1. We establish a small 'Base Set' of known-non-matching reference Writers that are compared against the test document.
  - In FLASH ID, this is the fixed column writers in the FLASH ID database.
  - In these experiments, we use a 50 Writer Base Set.
2. The required Rewards infrastructure for computing an Objective Measure of Similarity is already established in FLASH ID.

We are able to construct an informative error distribution that is predictive of the rarity of the measured similarity of any reference Writer of interest to the test document.

## 2.0: Summary of Findings

### 2.1: Basic Principles for Scoring Fingerprints and Handwriting

Figure 6 is a histogram of the Standardized t-Test Data (i.e.,  $z_{i,k}^j$  scores) for the known non-mate writers from the project. The 967,750 scores in the histogram comprise 2,450 scores for each of the 395 known non-mate writers. For each writer, the 2,450 scores will be reduced via hierarchical computation to a single Objective Similarity Statistic.

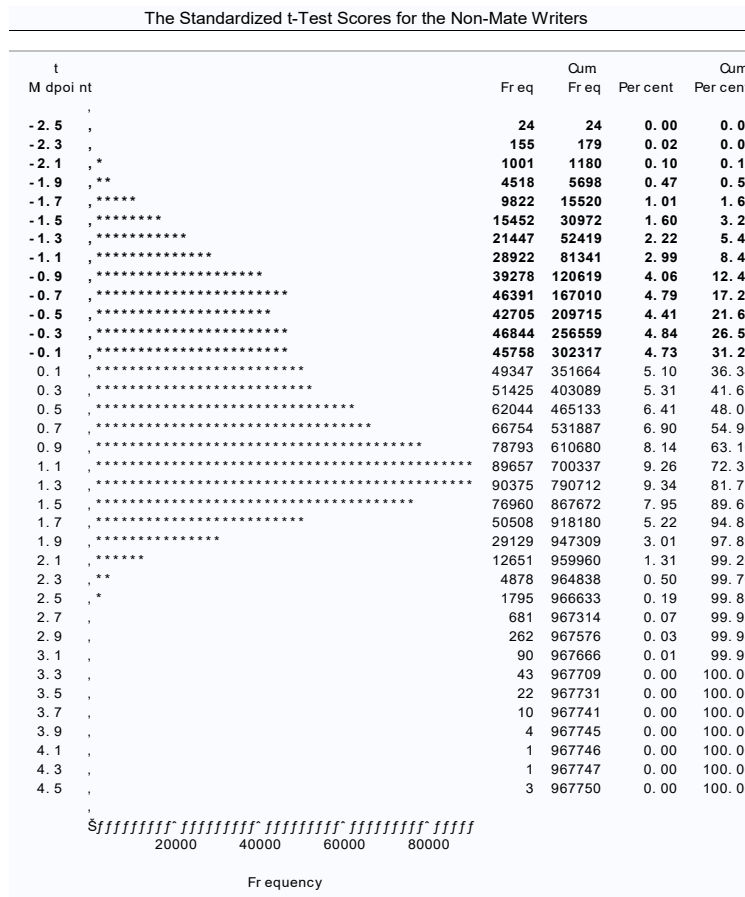


Figure 6: Histogram of the Standardized t-Test Data (i.e.,  $z_{i,k}^j$  scores) for the known non-mate writers from the FLASH ID feasibility task.

As is the case with latent fingerprint analysis, it is necessary to isolate the informative left-tail of the data in Figure 6. The left-tail towards negative scores is informative in that it is the direction of increasing similarity of the writers' known writings to the test document.

The Negative Similarity Scores are highlighted in the Figure 6 histogram. Some details of the informative left-tail are presented by the two plots in Figure 7. Each plot presents data for a single Base Set Writer  $j$ . Note that each of the 2,450 Standardized t-Test Scores is indexed by 3 letters  $(i,j,k)$ :  $k$  is the known non-mate reference writer;  $i$  and  $j$  are Base Set writers. In the scoring structure, Writers  $k$  and  $i$  are compared relative to how each competes with Writer  $j$  for similarity to the test document. Note that the 'True Writer' data outliers the true non-writer data.

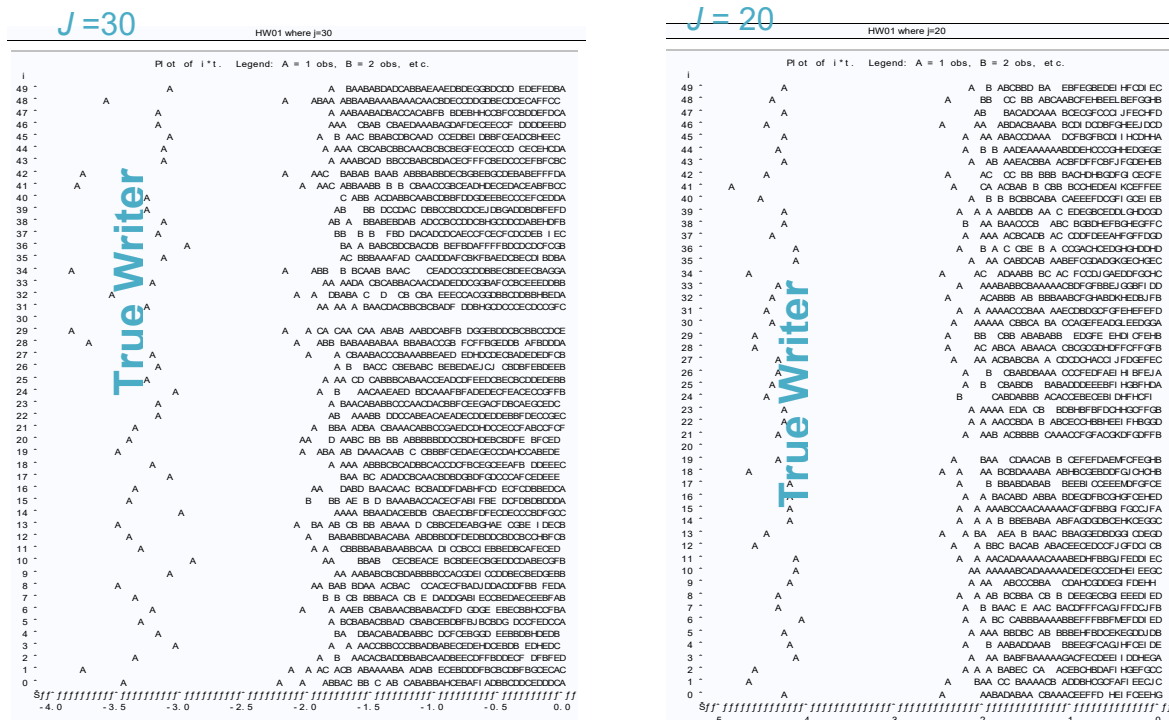


Figure 7: Sample strips of data for a single Base Set writer  $j$  showing outlying scores (left).

Each of the above plots (Figure 7) has strips of data for a single Base Set Writer  $i$ . The outlying scores to the left on each line indicate the True-Writer score. The first step in computing the Objective Similarity Score is to take the Median, for a fixed Base Set Writer  $j$ , of the  $(i,j,k)$

scores for each Writer  $k$ . Two plots in Figure 8 display the first level median data for the FLASH ID Feasibility Task.

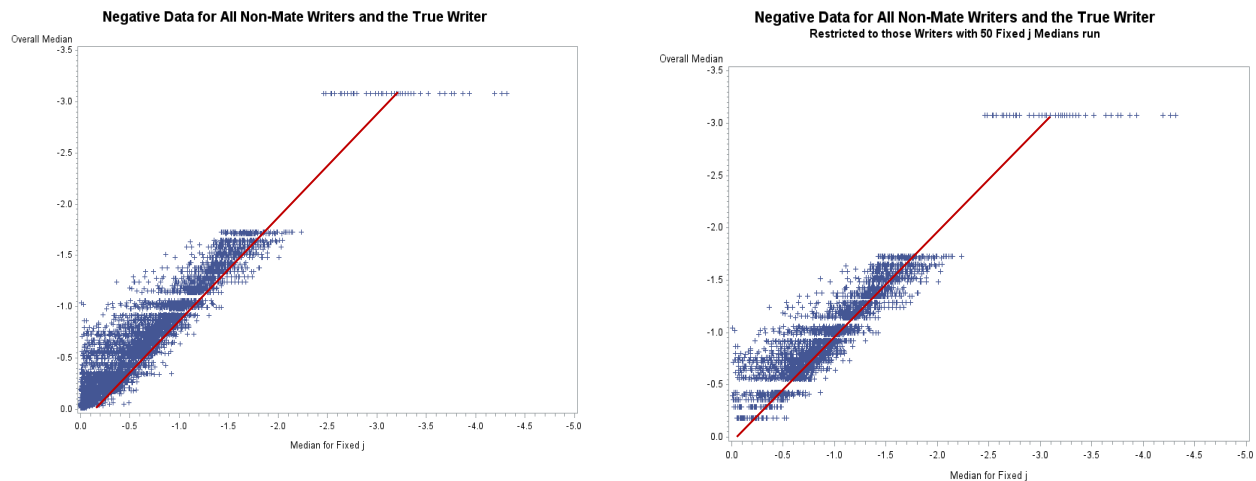


Figure 8: First level median data is plotted at the level (on the vertical axis) of the Overall Hierarchical Median for writer  $k$ . A red line marks the Overall Hierarchical Median.

The plots in Figure 8 show Overall Hierarchical Median for Writer  $k$ . A red line marks the Overall Hierarchical Median. The plot on the right has the negative Standardized t-Test data just for those Writers  $k$  that have all 50 of their first level median data negative.

Observe from the two plots in Figure 8 that the plot on the right is the more extreme right tail of the data in the left side plot. It is the data in the right-side plot that is statistically most informative concerning similarity to the test document, and it is the data that we use to model non-mate similarity to the test document.

The histogram shown in Figure 9 presents the 78 overall medians for the data from the right side plot above together with the corresponding overall median from the True-Mate Writer. These 78 non-mate data values will be the basis for modeling non-mate similarity to the test document.

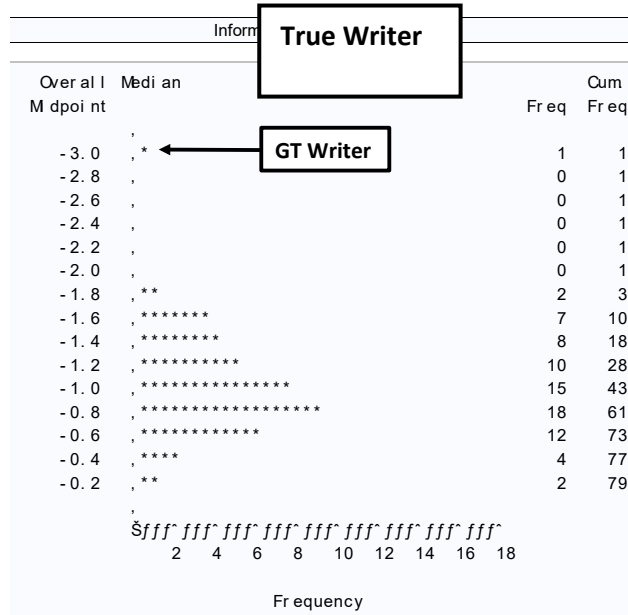


Figure 9: 78 overall medians for the data from the Figure 8 right-side plot together with the corresponding overall median from the True-Mate Writer.

## 2.2: Analysis with HPSEVERTY Procedure (SAS System)

The SAS System HPSEVERTY procedure is used to evaluate the fit of various probability distributions to the non-mate overall medians. The candidate distributions are all defined for positive values. Therefore, for modeling purposes, the negative scores in the above histogram are treated as if they were positive values. Table 1 presents the results from the HPSEVERTY procedure. The selected fitted distribution is the Weibull.



Table 1: Summary of Results from HPSEVERTY Procedure

Model Selection			
Distribution	Converged	AICC	Selected
Burr	Maybe	68.02523	No
Exp	Yes	154.67615	No
Gamma	Yes	69.49253	No
Igauss	Yes	79.99473	No
Logn	Yes	77.13266	No
Pareto	Yes	156.82958	No
Gpd	Yes	156.78352	No
Weibull	Yes	<b>65.85149</b>	<b>Yes</b>

- Weibull =  $(1/1.0501) * \text{Median} - .0442/1.0501$
- The median score for the True Writer is -3.07965. So, the corresponding Weibull Score for the True Writer is 2.7916 [Accounting for sign changes].
- The probability of a Weibull Score being at least as large as 2.7916 is:  $1 - 0.9999999961 = 3.9087E-9$ .
- The corresponding Rarity is 1/255,836,483.

The quantile-quantile plot for the Weibull distribution fit is presented below in Figure 10. This plot suggests a good fit of the data to Weibull distribution

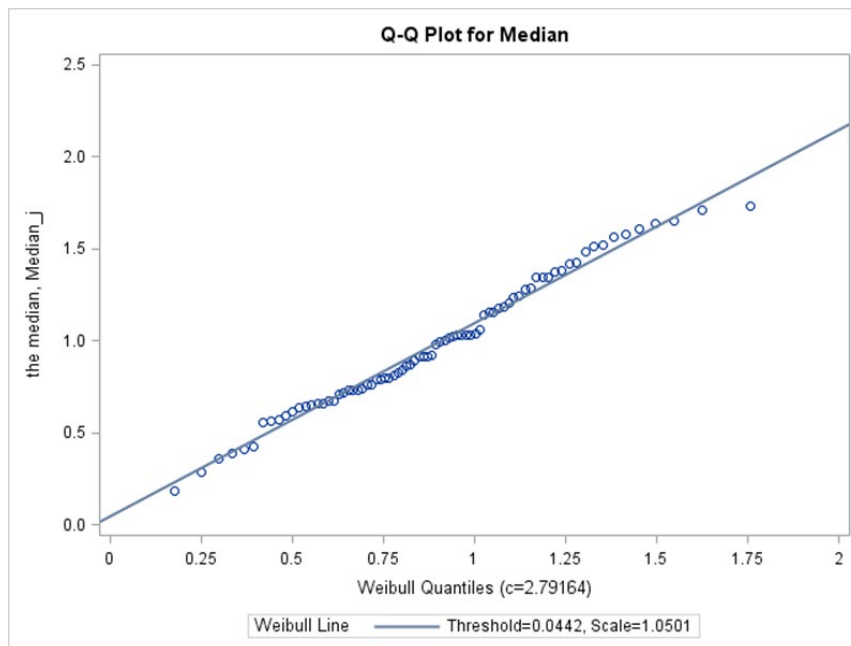


Figure 10: Quantile-Quantile Plot for the Weibull Distribution Fit.

## 2.3: Deriving the Objective Score

The data analysis and model fitting just presented suggests that the general approach for designing an objective score used in the NIJ Grant for latent fingerprint scoring can be applied to handwriting. The general approach used in the NIJ Grant is:

- Reduce the data to the tail of data that is informative concerning the competitive similarity of the non-mate Writers to the test sample.
- Fit a Normal Probability Distribution to the informative tail of the data.

There is a significant difference between the similarity data in the latent fingerprint case and the handwriting case. The data in the latent fingerprint case is based on locations dense across the Level 2 structure of the latent image. The data in the handwriting case is based on discrete segmentation of handwriting. Accordingly, the families of probability distributions that are appropriate for modeling non-mate similarity differs between the two cases.

The approach presented above for handwriting is now applied to one of the latent fingerprint examples featured in the NIJ Grant Final Research Report. That example uses the *Ugly* U260 Latent from the NIST Special Database 27.

The U260 data used here consists of a random sample of 19,809 known non-mate fingerprint images from a very large fingerprint image data base. The corresponding Standardized t-Test Data consists of 48,532,042 scores that are presented in the histogram on the left side of Figure 11. The intervals for negative scores in this histogram are highlighted in blue. The 794,530 first level negative medians (i.e., medians of the negative Standardized t-Test scores for a fixed Base Set Image  $j$ ) are presented in the histogram on the right side of Figure 14.

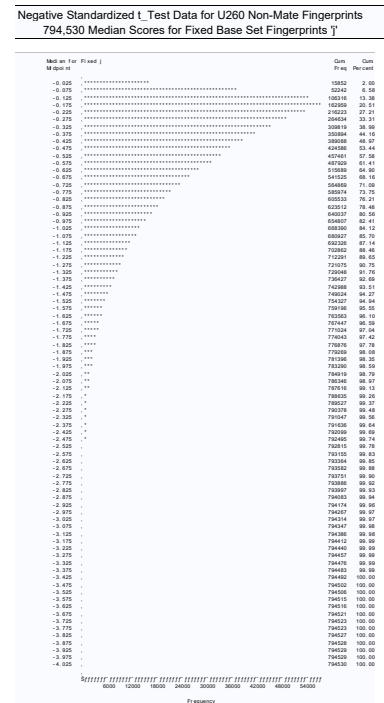
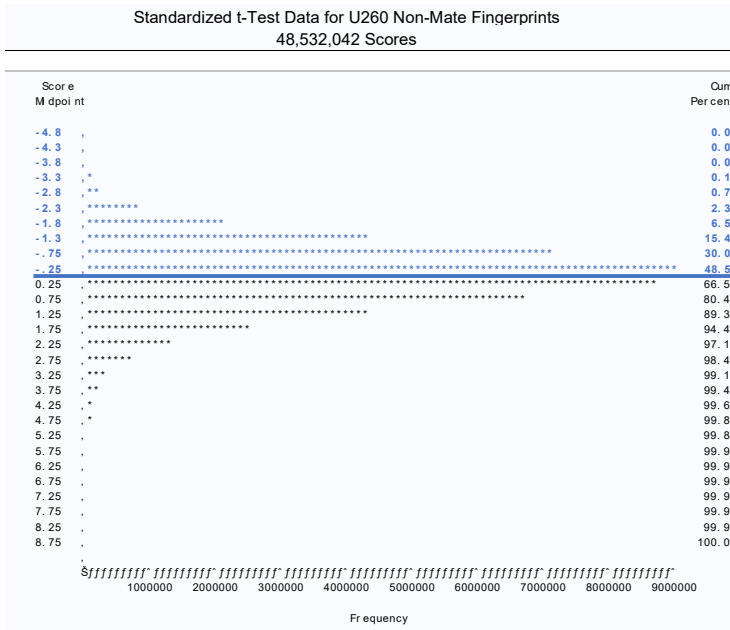


Figure 11: Latent Fingerprint Distributions from NIJ Grant.

We can demonstrate the application of the handwriting analysis to the U260 latent fingerprint data. Steps used for both handwriting and fingerprints as illustrated by the U260 example:

1. Cut the first level hierarchical medians (within one Base Set j) to the negative ones.
2. Cut the overall median data to those based on 50 first level medians.
3. Fit a distribution from the family of distributions for positive data. [We flip the negative overall median data to do this.]

The two plots in Figure 12 present data analogous to the two similar looking plots of the first level median data presented above for the current handwriting project.

In each of these plots, the first level Median data is plotted at the level (on the vertical axis) of the overall hierarchical median for fingerprint image  $k$ . A red line marks the overall hierarchical median. The plot on the left has all of the negative Standardized t-Test data. The plot on the right has the negative Standardized t-Test data just for those fingerprints  $k$  that have all 50 of their first level median data negative. Observe from the two plots that the plot on the right is the more extreme right tail of the data in the left side plot. It is the data in the right side plot that is statistically most informative concerning similarity to the Latent Image, and it is the data that we use to model non-mate similarity to the Latent Image to stay consistent with the modeling of the current handwriting project.

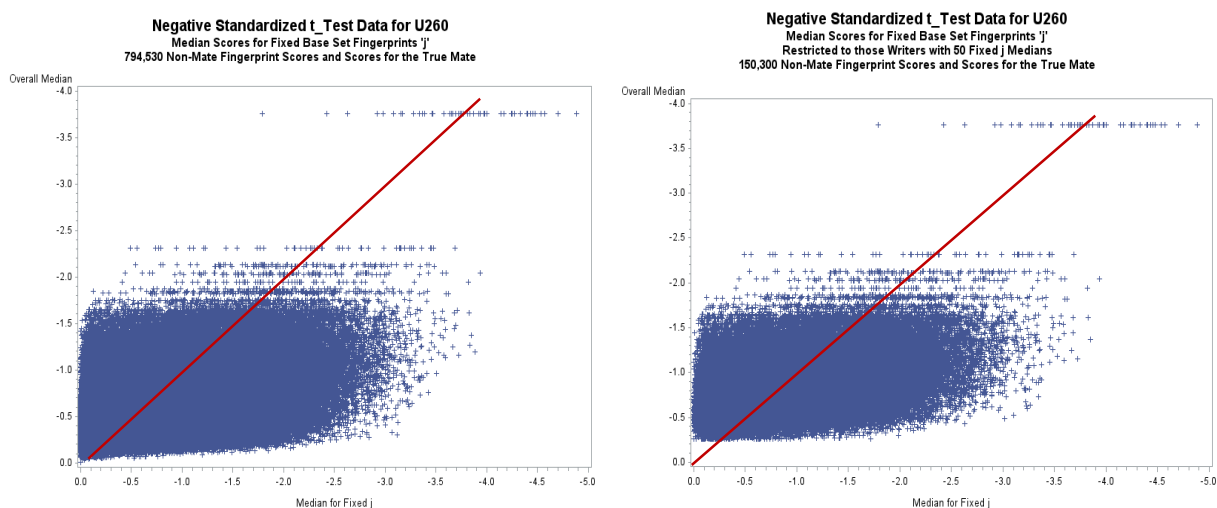


Figure 12: First level median data is plotted at the level (on the vertical axis) of the Overall Hierarchical Median for fingerprint image

The histogram in Figure 13 presents the 3,007 overall medians for the data from the right side plot above together with the corresponding overall median from the True-Mate Fingerprint Image. These 3,007 non-mate data values will be the basis for modeling non-mate similarity to the Latent Image.

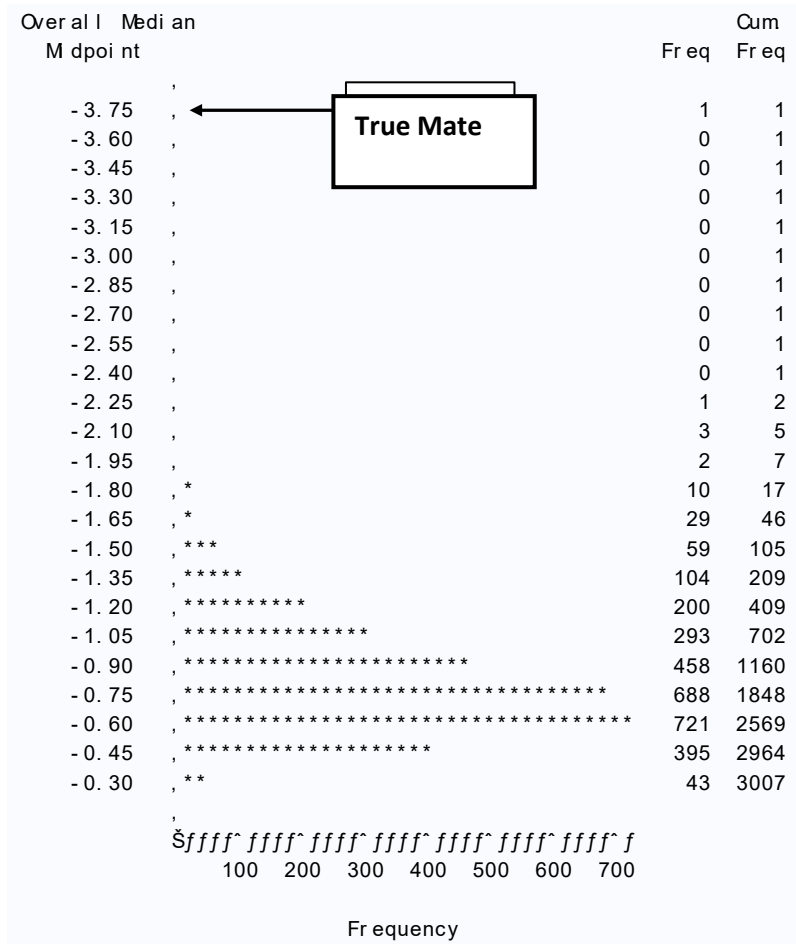


Figure 13: Chart showing 3,007 Overall Medians for the data from the right side plot from Figure 12 together with the corresponding Overall Median from the True-Mate Fingerprint Image

The SAS System HPSEVERTY procedure is used to evaluate the fit of various probability distributions to the non-mate overall medians. The candidate distributions are all defined for positive values and are the same distributions used for the current handwriting project. Again, for modeling purposes, the negative scores in the above histogram are treated as if they were positive values. Applying the HPSEVERTY procedure, the selected fitted distribution is the Lognormal.

- Lognormal =  $(1/0.8878078) * \text{median} + .119/0.8878078$ .
- The median score for the True-Mate Fingerprint Image is -3.7965.
- The corresponding Lognormal Score [accounting for sign changes] for the True-Mate Image is 4.36801.
- The corresponding Rarity is 1/88,357,999. This value is much less Rare than that computed using all of the informative data in the latent fingerprint case. The purpose here has been to illustrate the portability of the technique used for handwriting to the latent fingerprint case.

## **2.4: Deriving the Objective Score for Handwriting**

Sciometrics replicated the objective scoring method developed for fingerprints to determine feasibility for handwriting. Specific steps included:

- We selected FBI 500 Writer 00011 as the test case writer for the feasibility analysis.
- We established a small 'Base Set' of known-non-matching references that are compared against the test document.

## **2.5: Other Considerations for Applying Objective Scoring to Handwriting**

The following items represent considerations arising from this feasibility report that remain relevant to future extension of the scoring method for fingerprints to handwriting.

# SCIOMETRICS

- **Running Time/Storage Requirements:** additional running time and storage required scales with size of known-non-matching database (equivalent to running an additional Nx50 linear database recognition).
- **Required Samples:** the data run was based on what was available from the FBI 500 (the largest same-language set we have available). Larger data sets *could* produce better results, but the total effect is unknown because we currently lack the data. Equivalent samples would have to be obtained for the different languages wanting to be supported (for both the base set and known-non-matching set).
- **Grapheme Count:** as the grapheme count is reduced, the predicted rarity will also be reduced. We currently do not know a concrete rule for when the number of graphemes cannot produce a valid rarity prediction.
- **Additional Statistical Research:** the proof of concept is not a robust, complete, turnkey solution for creating these models in handwriting. Work is still required to tune the handwriting domain to work optimally for this objective scoring approach.
- **Large Scale Tests on Varied Data:** testing needs to occur at a larger scale with varied test and reference data sets across languages, database sizes, grapheme counts, etc.
- **FLASH ID Integration:** the code for this PoC is piecemeal and not all implemented in .NET. Test code is currently producing data that is being run through SAS procedures in an R&D environment. This will have to be translated fully into a .NET component integrated into FLASH ID.

## 3.0 APPENDIX: Examples of rarity results from Fingerprint Study

### Latents Fully Analyzed in the Grant

Latent	Median Observations for Analysis	Minutia	Remainder Percentage	GT Score	GT p-Value	Predicted Rarity	Log Base 10 of Rarity
B117	19,829	11	78	-7.01226	1.17E-12	852,912,259,109	11.93
B122	19,830	6	99	-4.65554	1.616E-6	618,926	5.79
B124	19,813	4	57	-4.39349	5.577E-6	179,295	5.25
B126	19,833	26	95	-8.44124	1.57E-17	63,694,444,445,493,800	16.80
B129	19,830	6	98	-1.93605	0.026431	37	1.57
B139	19,815	6	100	-5.26811	6.89E-8	14,509,698	7.16
B154	19,815	5	88	-5.11524	1.57E-7	6,382,703	6.81
B167	19,821	11	82	-6.84507	3.82E-12	261,641,392,900	11.42
B189	19,824	7	73	-4.45606	4.174E-6	239,582	5.38
B190	19,830	7	100	-5.11699	1.55E-7	6,442,390	6.81
B198	19,831	6	100	-4.15965	0.000015937	62,748	4.80
G012	19,823	9	100	-5.97732	1.13E-9	881,681,213	8.95
G073	19,827	19	70	-6.43406	6.21E-11	16,098,151,627	10.21
G087	19,824	12	100	-6.46668	5.01E-11	19,964,013,984	10.30
G095	19,831	12	78	-5.98839	1.06E-9	943,696,512	8.97
U202	19,816	4	100	-4.24494	0.000010933	91,468	4.96
U204	19,825	11	81	-6.73862	7.99E-12	125,079,736,825	11.10
U205	19,825	10	97	-7.61872	1.28E-14	78,064,662,183,219	13.89
U206	19,825	13	50	-6.04262	7.58E-10	1,318,974,409	9.12
U213	19,823	5	100	-5.21655	9.11E-8	10,971,549	7.04
U229	19,816	6	86	-5.78794	3.56E-9	280,682,634	8.45
U230	19,829	6	76	-5.25991	7.2E-8	13,877,093	7.14
U243	19,826	5	100	-3.21835	0.000644643	1,551	3.19
U246	19,546	4	88	-3.85782	0.000057201	17,482	4.24
U254	19,813	4	97	-3.99161	0.000032814	30,475	4.48
U260A	18,753	7	80	-7.05537	8.61E-13	1,161,864,294,637	12.07
U260B	19,775	7	76	-7.13324	4.9E-13	2,040,100,558,257	12.31
U281	19,819	11	93	-5.40428	3.25E-8	30,736,135	7.49
U291	17,289	5	86	-4.56591	2.487E-6	402,148	5.60
WHORL	19,825	15	99	-6.56278	2.64E-11	37,867,719,679	10.58

U260A is the common Base Set analysis  
 U260B is the 'Best 50' Base Set analysis