J Forensic Sci, Jan. 2005, Vol. 50, No. 1
Paper ID JFS2004040
Available online at: www.astm.org

Chi-keung Li,[1] Ph.D.; Nai-leung Poon,[1] Ph.D.; Wing-kam Fung,[2] Ph.D.; and Chi-ting Yang,[2] M.Phil.

Individuality of Handwritten Arabic Numerals in Local Population*

ABSTRACT: This paper reports the statistical study of writing habits for Arabic Numerals of 187 subjects in Hong Kong. A classification system of writing habits for Arabic numerals based on assigned codes of characteristic features such as slant, writing direction, relative position of strokes, angularity of turnings, shape of initial and ending strokes, etc. was developed. A set of characteristic codes representing the profile of writing habits pertaining to Arabic numerals was assigned to each writer. Apart from the distribution of characteristic features, statistical analysis of the assigned codes demonstrated homogeneity of individual hand-writing patterns. It has been shown that irrespective of the structural simplicity of Arabic numerals, no two individuals exhibited the same set of characteristic codes. The findings support the hypothesis of individuality in handwriting.

KEYWORDS: forensic science, writing habits, Arabic numerals, statistical study, classification system, individuality in handwriting

Handwriting identification is based on the hypothesis that natural handwriting is the result of an individual's physical and mental endowment as influenced by surroundings, condition, system and methods of instruction under which it is acquired. Writing habit, being a product of long-term adaptation to the needs and abilities of the writer, is believed to be unique. Various classification systems for handwriting have been suggested in the past, and reviews on this by Huber and Headrick (1), Osborn (2), and Hilton (3,4) suggest that a particular writer may be identified from writing attributes such as form, direction, slant, and proportion. Fundamental differences in either form or movement are the means of distinguishing numerals written by different individuals. Strach (5) reported a system for the classification of handwritten numerals. Recently, Srihari et al. (6) reported the use of computer algorithms for extracting features from scanned images of handwriting. Using computer technology, many extracted handwriting features are amenable to statistical analysis, large-scale comparisons between handwriting specimens from different persons can be accomplished, and the likelihood of chance-match can be assessed.

This paper reports a statistical study on writing habits for Arabic numerals of 187 subjects in Hong Kong. It is hoped that results of the analysis can be used to demonstrate the homogeneity of subjects in their handwriting patterns and that the hypothesis of individuality of handwriting could be verified. To begin with our study, Arabic numerals were chosen because they are structurally simple, unambiguous and can be easily distinguished macroscopically. Hierarchical cluster analysis and pair-wise comparison between individuals were adopted in the study to measure the individuality of handwriting pattern.

[1] Questioned Documents Section, Government Laboratory, Ho Man Tin Government Laboratory, 88 Chung Hau Street, Kowloon, Hong Kong Special Administrative Region, People's Republic of China.

[2] Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region, People's Republic of China.

* The preliminary result of this work has been circulated in the 2003 Annual Meeting of the American Society of Questioned Document Examiners.

Received 8 February 2004; and in revised form 2 September 2004; accepted 2 September 2004; published 8 Dec. 2004.

Materials and Methods

A questionnaire was designed for the collection of numeral specimens. 187 subjects in the local population were asked to follow the instructions on the questionnaire: to write Arabic numerals five times as in their normal course of business, using their own pens and their accustomed hands. The background details of the subjects are shown in Table 1.

Characteristic features such as slant, writing direction, relative position of strokes, angularity of turnings, shape of initial and ending strokes, etc. from the ten numerical digits "0" to "9" were selected, and a code was then assigned to each characteristic feature as listed in Table 2, and Figs. 1–3 illustrate the assignment of characteristic codes for numerals "0", "1", and "5". Specimens of Arabic numerals were examined microscopically, using a Nikon SMZ-2B microscope. Data analysis of the writing features based on the assigned classification codes was carried out using the method of hierarchical cluster analysis. In this study, the classification codes for a particular feature of selected numerals were re-coded to a number of binary variables in which "1" refers to the presence of the code and "0" refers to its counterpart (i.e., absence). Proximity

TABLE 1—*Personal details of the 187 subjects.*

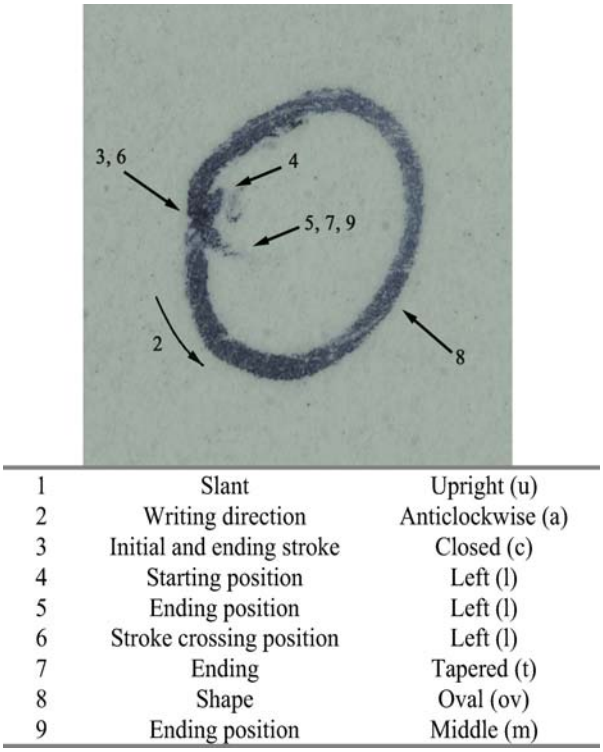
Age	<20	4%
	20–29	39%
	30–39	32%
	40–49	23%
	>50	2%
Education Background	Primary	5%
	High School	78%
	University	17%
Writing Frequency	Always	84%
	Sometimes	13%
	Seldom	3%
Location of Education	Hong Kong	96%
	China	2%
	Foreign	2%
Handedness	Right	99%
	Left	1%

TABLE 2—*Assigned codes of writing features and their distribution of occurrence exhibited by 187 subjects.*

Numeral	Features	Code	% Occurrence	Code	% Occurrence	Code	% Occurrence	Code	% Occurrence
0	Slant	Forward (f)	52.7%	Backward (b)	1.6%	Upright (u)	45.7%		
	Writing direction	Clockwise (c)	2.2%	Anti-clockwise (a)	97.8%				
	Initial and ending stroke	Open (o)	10.3%	Close (c)	89.7%				
	Starting position	Left (l)	54.8%	Middle (m)	42.5%	Right (r)	2.7%		
	Ending position	Left (l)	55.9%	Middle (m)	34.9%	Right (r)	9.2%		
	Stroke crossing position	Left (l)	58.4%	Middle (m)	32.4%	Right (r)	1.6%	No (n)	7.6%
	Ending position	Tapering (t)	82.8%	Blunt (b)	17.2%				
	Shape	Oval (ov)	72.0%	Elongated (e)	27.4%	Flatten (f)	0.6%		
	Ending position	Upper half (u)	74.2%	Lower half (l)	20.4%	Middle (m)	5.4%		
1	Slant	Forward (f)	68.6%	Backward (b)	9.8%	Vertical (v)	21.6%		
	Initial hook	Absence (a)	81.3%	Left (l)	4.8%	Right (r)	13.9%		
	Serif	Absence (a)	100.0%	Presence (p)	0.0%				
	Ending position	Blunt (b)	10.2%	Hook (h)	25.3%	Tapering (t)	64.5%		
2	Slant	Forward (f)	11.2%	Backward (b)	39.6%	Upright (u)	49.2%		
	Initial stroke direction	Downward (d)	58.3%	Horizontal (h)	17.6%	Upward (u)	24.1%		
	Ending stroke direction	Downward (d)	17.1%	Horizontal (h)	23.0%	Upward (u)	59.9%		
	Connection at the bottom	Open (o)	50.3%	Loop (l)	23.0%	Retrace (r)	26.7%		
	Ending	Tapered (t)	93.6%	Blunt (b)	6.4%				
	Turning at the top	Angular (a)	9.6%	Round (r)	90.4%				
	Top/Bottom ratio	Top (t)	0.5%	Bottom (b)	82.4%	Similar (sim)	17.1%		
3	Slant	Forward (f)	18.7%	Backward (b)	33.2%	Vertical (v)	48.1%		
	Initial stroke direction	Downward (d)	70.6%	Horizontal (h)	16.0%	Upward (u)	13.4%		
	Upper turning	Angular (a)	6.4%	Round (r)	93.6%				
	Lower turning	Angular (a)	12.3%	Round (r)	87.7%				
	Ending	Tapered (t)	78.6%	Blunt (b)	21.4%				
	Ending portion	Hook (h)	8.6%	Loop (l)	4.4%	Simple (s)	87.0%		
	Size of upper/lower	Lower Larger (l)	42.7%	Upper larger (u)	8.1%	Similar (s)	49.2%		
	Upper/lower relationship	Retrace (r)	13.9%	Simple (s)	86.1%				
4	Turning to the left	Angular (a)	25.1%	Round (r)	74.9%				
	Loop on the left	Yes (y)	0.0%	No (n)	100.0%				
	Connection between horizontal and vertical strokes	Loop (l)	12.3%	Open (o)	87.7%				
	Relation between slanting and vertical strokes	Close (c)	5.4%	Open (o)	94.6%				
	Ratio of vertical stroke above & below the horizontal stroke a/b	Longer (lo)	50.3%	Shorter (sh)	9.6%	Similar (si)	40.1%		
	Top part of vertical stroke relative to left slanting stroke	Taller (ta)	6.5%	Shorter (sh)	46.2%	Similar (si)	47.3%		
	Left slanting stroke (a)/ portion of vertical stroke below horizontal stroke (b)	Longer (lo)	57.8%	Shorter (sh)	7.4%	Similar (si)	34.8%		
	Ending of vertical stroke	Tapered (t)	91.4%	Blunt (b)	8.6%				
5	Slant	Forward (f)	54.5%	Backward (b)	6.4%	Upright (u)	39.1%		
	Position of starting of Horizontal stroke related to vertical stroke	Above (a)	5.9%	Below (b)	63.1%	Similar (s)	31.0%		
	Crossing of Horizontal stroke/curve	Yes (y)	19.8%	No (n)	31.5%	Touch (t)	48.7%		
	Position of Crossing	Top (t)	34.8%	Middle (m)	23.5%	Lower (l)	15.0%	No (n)	26.7%
	Ending Orientation	Upward (u)	43.9%	Downward (d)	25.1%	Horizontal (h)	31.0%		
	Ending	Tapered (t)	73.8%	Blunt (b)	2.7%	Loop (l)	5.3%	Hook (h)	18.2%
	Bottom portion	Hook (h)	5.9%	Straight & Angular (sa)	44.4%	Round (r)	49.7%		
	Ratio	Top (t)	48.7%	Lower (l)	4.3%	Equal (e)	47.0%		
	Turning stroke	Round (r)	63.6%	Angular (a)	20.9%	Nil (n)	15.5%		
6	Slant	Forward (f)	69.0%	Backward (b)	1.1%	Upright (u)	29.9%		
	Initial hook	Presence (p)	35.3%	Absence (a)	64.7%				
	Ending	Cross (c)	68.4%	Open (o)	31.6%				
	Shape of loop	Flat (f)	48.4%	Round (r)	36.4%	Oblong (o)	15.2%		
7	Slant	Forward (f)	86.6%	Backward (b)	3.2%	Upright (u)	10.2%		
	Turning	Round (r)	41.2%	Angular (a)	58.8%				
	Stroke initial	Presence (p)	17.1%	Absence (a)	82.9%				
	Crossing bar	Presence (p)	26.2%	Absence (a)	73.8%				
	Ending	Tapered (t)	58.8%	Blunt (b)	14.5%	Hook (h)	26.7%		
	Vertical portion	Straight (s)	50.8%	Curved (c)	49.2%				
	Horizontal/Vertical	Horizontal Longer(lo)	1.1%	Horizontal shorter (sh)	87.2%	Similar (sim)	11.7%		
	Crossing position	Upper (u)	7.5%	Middle (m)	18.7%	Absence (a)	73.8%		
	Turning	>90 (b)	7.9%	<90 (s)	92.1%				

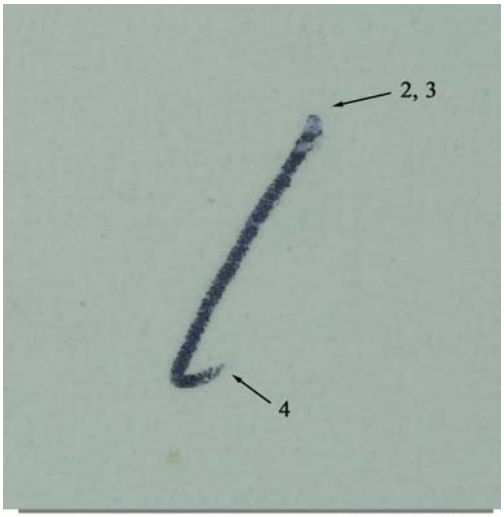
TABLE 2—*Continued.*

Numeral	Features	Code	% Occurrence	Code	% Occurrence	Code	% Occurrence	Code	% Occurrence
8	Slant	Forward (f)	61.5%	Backward (b)	2.7%	Upright (u)	35.8%		
	Ending	Tapered (t)	81.8%	Blunt (b)	17.6%	Hook (h)	0.6%		
	Starting loop writing movement	Anticlockwise (a)	91.4%	Clockwise (c)	8.6%				
	Starting position	Left (l)	27.3%	Middle (m)	59.9%	Right (r)	12.8%		
	Ending position	Left (l)	9.1%	Middle (m)	31.0%	Right (r)	59.9%		
	Initial stroke crossing	Yes (y)	31.0%	No (n)	69.0%				
	Ending stroke crossing	Yes (y)	84.0%	No (n)	16.0%				
	Special shape	E shape (e)	1.6%	2 Circles (2)	2.7%	Ordinary (o)	95.7%		
	Extension after lower cross	No (n)	24.1%	Minimal (m)	25.1%	Obvious (o)	50.8%		
	Separation between upper & lower loop	No (n)	76.5%	Minimal (m)	18.7%	Obvious (o)	4.8%		
	Initial portion	No (n)	65.8%	Minimal (m)	21.9%	Obvious (o)	12.3%		
	Upper/lower relationship	Upper larger (u)	15.2%	Lower larger (l)	15.8%	Similar (s)	69.0%		
9	Slant	Forward (f)	77.0%	Backward (b)	5.3%	Upright (u)	17.7%		
	Ending	Tapered (t)	74.1%	Blunt (b)	19.4%	Hook (h)	6.5%		
	Writing movement	Anticlockwise (a)	98.9%	Clockwise (c)	1.1%				
	Position of loop crossing	Loop (l)	34.4%	Vertical (v)	7.5%	Open (o)	18.3%	Touch (t)	39.8%
	Starting position of loop relate to vertical	Left (l)	71.7%	Right (r)	5.8%	Middle (m)	22.5%		
	Upper turning	Curve (c)	46.0%	Angular (a)	47.0%	Round (r)	7.0%		
	Position of turning relate to starting	Higher (h)	13.9%	Lower (l)	31.0%	Similar (s)	43.3%	No turning (n)	11.8%
	Vertical stroke	Straight (s)	57.2%	Curve (c)	42.8%				
	Loop length to total height	1/2 (2)	10.2%	1/3 (3)	77.0%	1/4 (4)	12.8%		
	Starting position of loop	Top (t)	3.2%	Upper half (u)	11.2%	Lower half (l)	85.6%		



1	Slant	Upright (u)
2	Writing direction	Anticlockwise (a)
3	Initial and ending stroke	Closed (c)
4	Starting position	Left (l)
5	Ending position	Left (l)
6	Stroke crossing position	Left (l)
7	Ending	Tapered (t)
8	Shape	Oval (ov)
9	Ending position	Middle (m)

FIG. 1—*The assignment of characteristic codes for numeral "0".*

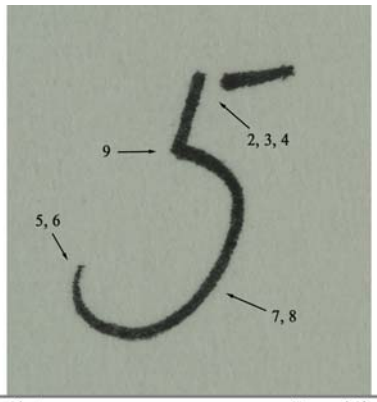


1	Slant	Forward (f)
2	Initial hook	Absence (a)
3	Serif	Absence (a)
4	Ending position	Hook (h)

FIG. 2—*The assignment of characteristic codes for numeral "1".*

of dissimilarity was generated for binary data that ranged from 0 to 1 (7–9). The fourfold table was then computed as bc/(n**2), where b and c refer to the diagonal cells corresponding to codes present on an object but absent on the other, and where n is the total number of subjects involved in the study (7–9). It is noted that hierarchical cluster analysis is a set of statistical techniques that is particularly useful for classifying a set of objects into constituent groups or clusters which minimize the variation between members of the same group without making assumptions about the number of groups or the group structure. The dendrogram or tree, using average linkage between subjects, was used for demonstrating the procedure of statistical classification. The dendrogram provides a convenient way of grouping subjects with similar writings of a numeral into clusters.

In order to quantify/demonstrate the individuality in handwriting pattern, pair comparison of 187 subjects was performed. A total of $187*186/2 = 17\,391$ pairs of comparison were computed for each Arabic numeral, and the number of pairs with exactly the same handwriting pattern in the selected characters was counted.



1	Slant	Forward (f)
2	Position of starting of Horizontal stroke related to vertical stroke	Similar (s)
3	Crossing of Horizontal stroke/slanting stroke	No (n)
4	Position of Crossing	No (n)
5	Ending Orientation	Upward (u)
6	Ending	Tapered (t)
7	Bottom portion	Hook (h)
8	Ratio of top and bottom	Lower emphasis (l)
9	Turning stroke	Angular (a)

FIG. 3—*The assignment of characteristic codes for numeral "5".*

Results and Discussion

In the present study, the adopted classification system for Arabic numerals was confined mainly to the pictorial forms of individual numerals and their associated writing movements. Other delicate writing features, which include pen pressure, writing skill, and line quality, were not considered in our study, as these last three features could not be assessed accurately and objectively. Writing specimens of isolated numerals used in our study also reduced the chance of any connecting strokes between numerals that might distort numerical forms or affect ending strokes. As indicated in Table 1, the majority of the subjects has an education background of high school or above (95%), and most of them are frequent writers (84%). Examination of the writing specimens from the 187 subjects revealed that most of the numerals displayed a forward slant. The writing direction of the loop containing numerals is mostly anti-clockwise with "round turnings." Other noticeable features were the round loop enclosures in most of the numerals "6", "8", and "9". Tapering ending strokes also were found in most of the numerals. The occurrence of tapering is in fact related to the motion of the writing instrument departing from the surface of the paper and is dependent upon the fluency and the speed of execution of a writer.

Document Examiners often encounter difficulties in the comparison of numerals, since the simple designs of the numerals limits the features that can be depicted. In the present study, a chain of easily assigned characteristic codes could be obtained from Arabic numerals written by 187 subjects. A typical example showing characteristic code patterns exhibited by two subjects is shown in Fig. 4. Table 2 presents the assigned codes of writing features of ten numerals ("0" to "9") and the corresponding percentage of occurrence for each code amongst the subjects considered. Taking numeral "1" as an instance of study, four characteristic features

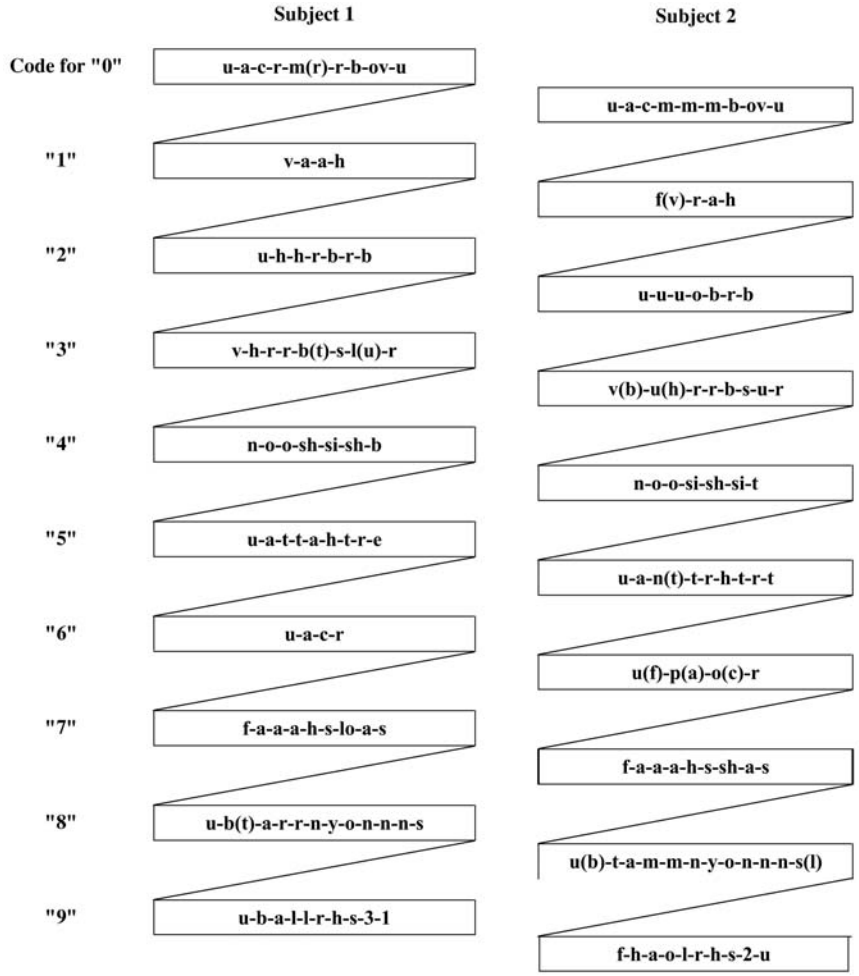


FIG. 4—*Characteristic code patterns exhibited by two subjects, n = 1 and n = 2 (see Table 2 for codes).*

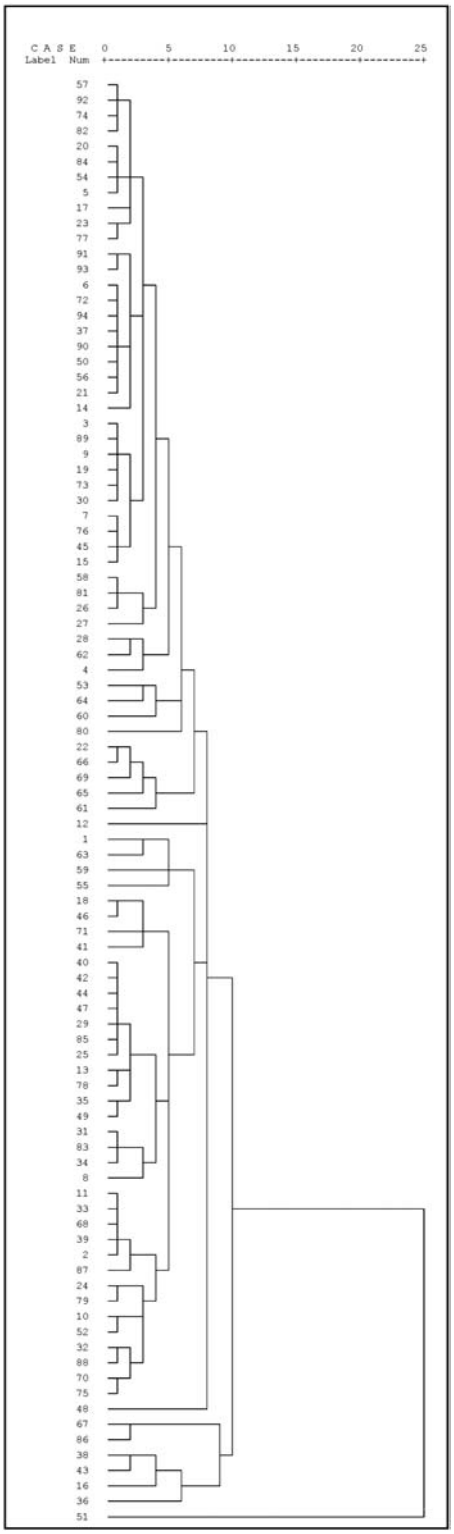


FIG. 5—*Dendrogram plot of hierarchical cluster analysis of numeral "0"*.

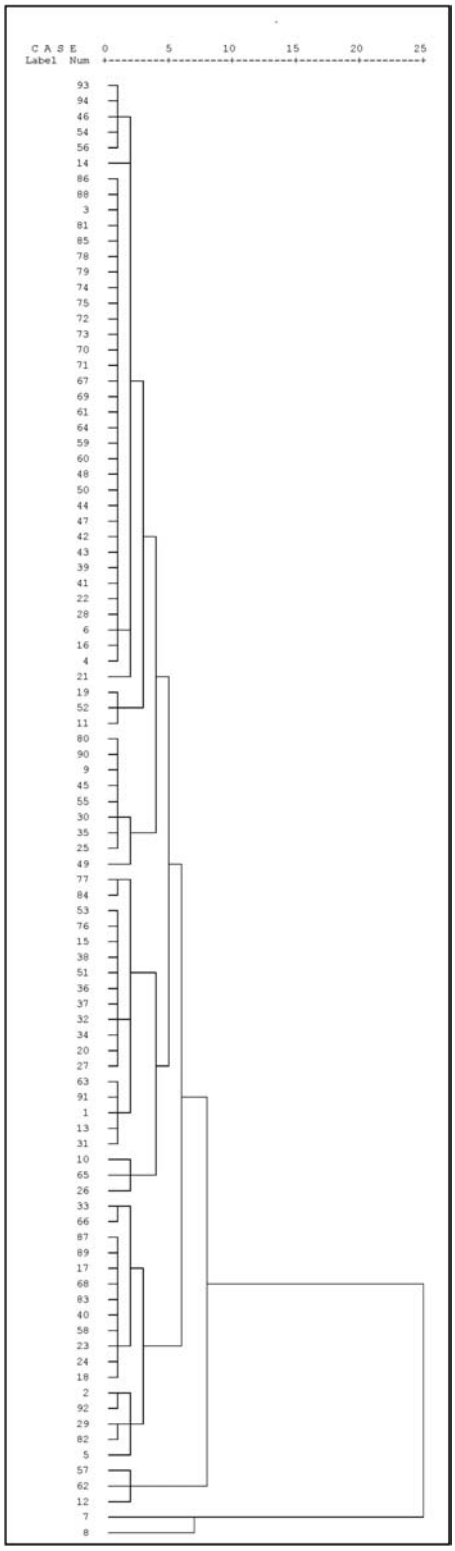


FIG. 6—*Dendrogram plot of hierarchical cluster analysis of numeral "1"*.

including slant, initial hook, serif, and its ending position can be selected. About 69% of the subjects wrote the numeral "1" in forward slant while only about 10% presented it in backward slant. About 80% and 100% of the subjects wrote the numeral "1" without initial hook and serif respectively. Only 10% of the subjects used a blunt ending stroke. In short, the most common writing features of numeral "1" are writing in forward slant without initial hook and serif and writing with a tapering end. Details of the breakdown of other numerals are also given in Table 2 for reference.

Figures 5–7 give a summary of the results of numerals "0", "1", and mixed "0-1", upon which hierarchical cluster analyses were performed. Numeral "0" is selected here because of its prevalent usage and the ease of assessment in any daily writing pieces (e.g., bank checks). In this work, numeral "1" is also adopted in view of its simple writing pattern amongst all other numerals. For clarity, only the odd subjects ($n = 94$) are selected as an illustration, where the selection is irrelevant to the variables of interest. Figure 5 depicts

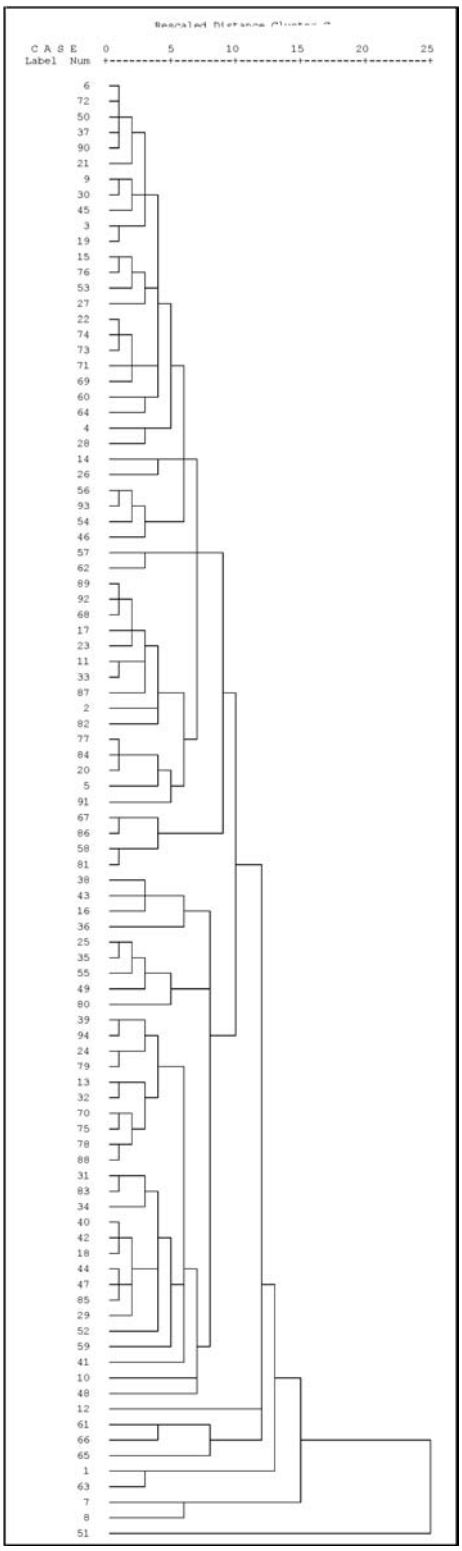


FIG. 7—*Dendrogram plot of hierarchical cluster analysis of numerals mixed "0-1".*

the classification plot of numeral "0". Apparently, subjects starting from individuals at the bottom of the dendrogram plot cluster closely together for individuals sharing common features and codes. Compared with numeral "0", a distinct pattern for numeral "1" is evident, and more subjects cluster together at the bottom of the dendrogram plot (Fig. 6). The analysis was performed again with a combined numeral "0-1" (Fig. 7). With more features and codes being considered, subjects show more heterogeneity from each other, and more cluster formations were observed at the bottom of the dendrogram plot. By considering two or more numerals, it becomes possible to perform more refined handwriting discrimination.

According to the hierarchical cluster analysis, numeral "1" embodies the simplest handwriting patterns where only four features and eleven codes are recorded. Numeral "5", consisting of nine assigned features, can be classified as the most informative numeral for distinguishing writers (results of cluster analysis for numerals other than "0" and "1" are skipped for brevity). The availability of different numerals increases the number of assigned features and codes, which in turn enhance the heterogeneity among subjects. Thus, combining two or more numerals reinforces dissimilarity among subjects and helps discriminate subjects from each other.

Table 3 records the presence of strictly homogeneous pairs in the selected features and codes of ten numerals. No single numeral could entirely discriminate all 187 subjects. Amongst all numerals considered, numeral "1" has the least variations of features, and 14.9% of the pair comparison have exactly the same writing patterns; the numeral "5" encompasses more distinctive features, and only 0.06% of the pairs have the same patterns. Thus for a single numeral, "5" is the most informative for the discrimination of handwritings. Numerals "8" and "9" also provide high discrimination power, as only 0.37% and 0.35% of the pairs were found to have exhibited no difference in their respective classification codes.

With the addition of a numeral in the pair comparison, the discrimination power is significantly enhanced. Some of these findings are also listed in Table 3 for reference. Numeral "0" again has been selected in the combination study due to its highly prevalent usage in our daily lives. The classification codes of a number of numerals in combination present a higher power of discrimination, with no combinations recording more than a dozen pairs of exactly the same patterns. However, a complete heterogeneity of writing patterns amongst subjects is reported for the mixed pair of "0-5", "0-8", and "0-9". According to the assigned features, the accumulated characteristics of only two Arabic numerals are sufficiently unique for the differentiation of the writers in the present study.

In subject-pair comparison, out of a total of 80 assigned writing features from the ten numerals, the most similar pair was found

TABLE 3—*Records of exactly the same patterns amongst period subjects in selected numerals.*

Single Numeral	No. of Pairs	No difference (%)	Combined Numerals	No. of Pairs	No difference (%)
0	73	0.4198	–	–	–
1	2599	14.9445	1-0	11	0.0633
2	220	1.2650	2-0	1	0.0058
3	257	1.4778	3-0	1	0.0058
4	532	3.0591	4-0	1	0.0058
5	11	0.0633	5-0	0	0.0000
6	632	3.6341	6-0	2	0.0115
7	324	1.8630	7-0	2	0.0115
8	65	0.3738	8-0	0	0.0000
9	61	0.3508	9-0	0	0.0000

TABLE 4—*Records of the maximum numbers of numerals with exactly the same patterns amongst the subject pairs.*

Subject pairs entry No.	Numerals match
141–143	1, 2, 6, 7
130–170	1, 4, 6, 8
24–122	1, 6, 7, 9

to share 63 writing features in common, and not a single pair of subjects was found to have displayed the same writing habit. Only three out of the possible 17 391 pairs were found to exhibit exactly the same set of writing patterns of four numerals (Table 4).

Conclusion

Pair comparison on the writing habits of Arabic numerals amongst 187 subjects has been performed in this study. Irrespective of the structural simplicity of numerals, the analysis was found to be effective in demonstrating the individuality of handwriting and, with the availability of only two different numerals, for discriminating the handwriting of two subjects. A much higher degree of discrimination could be found in considering more Arabic numerals.

Acknowledgments

The authors would like to record here their gratitude to Dr. D.G. Clarke, the Government Chemist of Hong Kong, for his interest and encouragement in this work and to Mr. S.C. Leung, Assistant Government Chemist for his valuable advices made on this work and for proofreading the paper. The work was partly supported by HKU CRCG research grant.

References

1. Huber RA, Headrick AM. Handwriting identification: facts and fundamentals. New York: CRC Press, 1999;152–64.
2. Osborn A. Questioned documents. 2nd ed. Chicago: Nelson-Hall Co., 1929;300–2.
3. Hilton O. Scientific examination of questioned documents. rev. ed. New York: Elsevier, 1982;217.
4. Hilton O. Identification of numerals. Int Crim Police Rev 1970;241: 245–50.
5. Strach SJ. Proposed research areas on handwriting comparison. Int J Forensic Doc Examiners 1998;4(4):310–2.
6. Srihari SN, Cha SH, Arora H, Lee S. Individuality of handwriting. J Forensic Sci 2002;47:1–17.
7. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: Wiley, 1990.
8. Everitt BS, Landau S, Leeseb M. Cluster analysis. 4th ed. New York: Oxford University Press, 2001.
9. SPSS Base 7.5 applications guide. Chicago: SPSS Inc., c1997.

Additional information and reprint requests:
Chi-keung LI, Ph.D.
Questioned Documents Section
Government Laboratory
Ho Man Tin Government Laboratory
88 Chung Hau Street, Kowloon
Hong Kong Special Administrative Region
People's Republic of China
E-mail: ckli@govtlab.gov.hk