Science and Justice xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

A statistical study of the relative difficulty of freehand simulation of form, proportion, and line quality in Arabic signatures

Abdulaziz Al-Musa Alkahtani, Andrew W.G. Platt *

Department of Forensic Science, Faculty of Science, Staffordshire University, College Road, Stoke-on-Trent, ST4 2DE, UK

ARTICLE INFO

Article history:
Received 27 April 2009
Received in revised form 13 August 2009
Accepted 24 August 2009
Available online xxxx

Keywords:
Arabic signatures
Simulation
Statistical analysis

ABSTRACT

This article compares the skill with which a large sample of Arabic writers was able to simulate different handwriting elements in two Arabic signatures. The data agree with the consensus of experts about Roman script: Form is significantly better simulated than Proportion, and Proportion is significantly better simulated than Line Quality. The fact that patterns long observed in Roman script simulation can be statistically demonstrated to exist in the simulation of Arabic signatures suggests that these may be widespread, if not universal, patterns of simulation among writing systems. This is also supported by similar results from a study of abstract characters (Leung et al. 1993), which suggests that these tendencies may even transcend linguistic writing systems. Among the Proportional elements are slant, alignment, spacing, and size.

1. Introduction

Consensus exists among forensic document examiners (FDEs) about the relative ease with which most handwriting elements can be simulated, but these assumptions had not been validated in controlled experiments. In recent decades, questions have been raised about the reliability of the assumptions underlying forensic handwriting analysis, owing to the lack of scientific testing of much of this knowledge. In 1993, the United States Supreme Court ruled that scientific evidence admitted in trials must be backed by scientific testing of the theories on which the techniques are based, error rates of the techniques, peer reviews of the tests, and acceptability in the relevant scientific community (Daubert v. Merrell Dow Pharmaceuticals 1993). Since that finding, U.S. courts have sometimes refused to admit handwriting analysis and other forensic methods, including fingerprint comparison, into evidence. These decisions are probably unfortunate from the viewpoint of justice, as critical evidence, based on techniques informally tested in long practice, may have been denied to jurors. But the current mood of skepticism offers an opportunity to forensic document examiners (FDEs) to test their assumptions and methods with controlled, statistical studies, and in the process to shed light on details on which there may have been disagreements. Thus, experimental examination of the assumptions and techniques of forensic document examination is an important and growing area of forensic document examination, to which the present research contributes.

This study applies statistical methods of comparison and probability estimation to one of the basic questions of forensic handwriting analysis: Which elements of handwriting are easiest and most difficult to simulate? The method used in this study was to elicit freehand simulation of Arabic signatures, under controlled conditions, by a large sample of Arabic speakers and writers, and then to compare the quality of the simulation. The phenomenon of freehand simulation is addressed in a majority of the texts and articles in forensic document examination, mainly because most forgeries are of this type. Another reason for this choice is that freehand simulation tends to produce errors in a great many elements, in contrast to traced simulation, which is detected chiefly by its characteristic line-quality errors. The simulation errors in a wide variety of elements are, of course, desirable in this experiment. The study of signatures is also the focus of most studies of simulation, and central to the field of forensic document examination, because signatures are involved in most forgeries. For that reason, signatures were chosen over other types of text in this experiment.

A unique aspect of this study is its focus on simulation of Arabic signatures by Arabic-writing participants. The literature on handwriting simulation deals almost entirely with simulation of Roman script, in which European languages are written, and predictions drawn from that literature are tested on Arabic script in this study. If the Roman script assumptions are supported by the Arabic script data, then we might hypothesize that the same assumptions are true for other scripts as well. Arabic writing, like much Western writing, is cursive, and both scripts are characterized by a fairly continuous line, broken between words. (There are six letters in Arabic that are also followed by mandatory breaks, even in the middle of words.) Mature, habitual writers of cursive Roman script produce a smooth Line Quality with unconscious rhythmic changes in pen pressure, all of which are lost when they carefully copy the shapes of another person's handwriting style [10,11,13,16,18,22]. It seems likely that

* Corresponding author.
E-mail addresses: almusa44@yahoo.com (A. Al-Musa Alkahtani), a.platt@staffs.ac.uk (A.W.G. Platt).

1355-0306/$ – see front matter
doi:10.1016/j.scijus.2009.08.003

this familiar characteristic of forgery in Roman script is also an important sign of forgery in Arabic script, and equally unnoticed by most simulators. It is possible that non-cursive scripts, such as Hebrew or Chinese, do not display the same subtle rhythm in Line Quality. However, Line Quality may still be important in these scripts; for instance, the habit of writing with especially strong pressure on downward or on horizontal strokes may differ from writer to writer, and may be largely unnoticed by simulators.

The signature elements studied. Questioned document examiners have divided handwriting elements into a variety of classification systems; Huber and Headrick [16] favor working with 21 basic elements that are useful to FDEs. In this study, three major elements were selected that have been suggested to differ in the ease with which they are simulated: Form, Proportion, and Line Quality. Proportion is further broken down into the elements of size, spacing, alignment, and slant, which have been suggested to differ in the ease with which they are simulated. In the forensic handwriting examination literature on Roman script, Form is generally considered the element most likely to be noticed by an amateur observer and most likely to be forged accurately. This may be, in part, because the forms of letters are consciously taught and learned in school, and also because of the relatively large size of letter shapes compared to more subtle aspects of writing. In this study, Form is taken to include the general letter shapes, the shapes of connections between letters, the treatment of dots, tendencies toward angular or toward curved strokes, and the shape of initial and terminal stroke shapes. In Roman script, a simulator who is unable to reproduce such noticeable features is likely to be a person who cannot write well (such as a child or an adult who is almost illiterate), has seen very little of the target writing, or who is simply careless [4,16,17,21]. Experimental evidence of the greater ease with which Form is noticed and simulated correctly is provided by Leung et al. [17] in a study of simulation of abstract symbols, which suggests that the advantage of Form in Roman script simulation may transcend specific writing systems.

The *Proportional elements of size, spacing, alignment, and slant* are thought by Roman script analysts to be less readily noticed and thus less accurately simulated than Form, and also likely to differ markedly in the writing of different individuals. Simulators are said to often unknowingly substitute their own writing habits for the target expressions of Proportional elements. In a study of simulation of abstract symbols [17], it was discovered that simulators who notice unusual Proportions tend to exaggerate them. For instance, an unusually long stroke might be simulated as even longer than the original. This discovery in simulation of symbols could well occur in the simulation of specific scripts, also. The painstaking measurements that are part of forensic handwriting examination tend to reveal small errors of Proportion that only the rare, equally painstaking simulator is likely to duplicate, and therefore we should expect such errors in simulation of Proportional elements to be detected by forensic document examiners.

Of the four Proportional elements investigated in this study, the element of size (of total signatures or other text) is regarded by Horan [15] as the most likely to be noticed and therefore simulated accurately, unless the forger does not know the usual size of the target's signature [9]. Nevertheless, errors in other aspects of size that are included in this study, such as relative sizes of different letters, or of heights and widths of letters, sizes of loops, and sizes of margins, are often mentioned as signs of forgery [9–11,13,16,18,21].

The slant of various strokes relative to the writing's baseline is regarded as a useful marker of forgery. Osborn [21] notes that the slant of different strokes is often highly consistent in individual handwriting, and thus only a few degrees of difference between slant in an original and in a forgery, if the forgery is also consistent, may be enough to identify fraud. Many people slant the upstrokes and down strokes of loops differently, which a simulator would not be likely to notice [10,11,13,16,18,21]. Thus, slant might be expected to be specially difficult to simulate well enough to deceive an FDE. Spacing between words and between letters within words varies greatly between individual styles. Letters may be spaced far apart, with long connections, and yet the words written by the same writer may be close together, or vice versa. Spacing is closely related to slant. That is, heavily slanting letters may allow little space between letters and also between words. In fact, slanting letters may overlap so much that there is actually negative space between them, and slanting beginning and ending letters of words may leave negative space between the words. Letters that are more upright may be more compact, with little space between them, unless they have unusually long connections [16,21].

Alignment of letters or of total handwriting relative to a ruled or imaginary horizontal baseline is described as highly characteristic of individual writers, but also largely unconscious, and therefore likely to be poorly simulated. Alignment is also taken to include habits of placement of signatures within the spaces available for them, which differ from person to person. In many cases, the alignment of writing gradually rises higher above the baseline over the course of the line, or less often, gradually falls farther below it. Some individuals align specific letters in special ways with relation to the baseline, and immature handwriting of children or very poorly educated adults may wander on both sides of the baseline [16,21].

Line Quality is widely regarded as the most difficult element of handwriting to accurately simulate, and thus the most useful indicator of forgery for the FDE. As cursive writing in Roman script becomes more mature, a writer gradually develops rhythmic alternations of pen pressure in the line which are often highly individual. Like similar rhythmic coordination in speaking, walking, and other largely habitual activities, the rhythm of writing is subconscious. It seems indubitable that this would be true of Arabic writing, as it is of Roman script writing, and in the author's experience it is indeed true. Only exceptionally sophisticated forgers are conscious of their Line Quality. When most forgers attempt to copy handwriting slowly and carefully, they lose their spontaneous, smooth line and produce the typical wavering, uncertain line, the *tremour de fraud*, that is a well-known red flag for forgery. The chief Line Quality signals of this self-conscious, careful drawing of another's handwriting are a wavering line with frequent changes of direction, pen lifts, covert (not large and open) patching and retouching, absence of alternating pressure patterns (or pressure patterns that differ from the original), and beginning and ending lines that begin or end abruptly, causing a blunt shape, as the pen is carefully placed before the drawing begins, rather than spontaneously tapering on or tapering off [8,10,11,13,16,18,21–23,25]. Simulation that is traced or copied is most likely to show poor Line Quality, whereas forgeries that are rapidly written in front of witnesses necessarily retain some of the forger's natural fluency. Sophisticated forgers often attempt spontaneous writing, capturing the general impression of the style, even though the Form and Proportion may actually contain many inaccuracies. The best forgeries are probably produced by a long period of tracing and copying until the target style becomes internalized and somewhat unconscious, gradually evolving into spontaneous freehand simulation without the target in view [2,12,16,18,19,21,25,26]. An uncertain, tremulous line may also be due to neuromuscular degeneration in old age or due to disease or effects of medicine at a younger age [1,3,5,7,14,16,20,21,24].

2. Methods and materials

The structure of the experiment. The aim of this experiment was to test the null hypothesis that the Line Quality, Form and Proportion elements discussed above do not differ significantly in their rankings for simulation quality. In order to obtain data on simulation quality, simulations of the same two Arabic signatures were elicited from 901 Arabic-writing participants. The two signatures were chosen to contain a maximum number of Arabic letters in a variety of word positions, where they might be written differently. The handwriting of the signatures was stylized to represent varied expressions of the six elements (Line Quality, Form and Proportion elements such as size, spacing, alignment and slant), in order to maximize the types of errors that could be made in simulating them.

The ranking of simulation quality. In order to compare the success of the signature simulations and inter-expert rating system was used. Each participant's simulations were independently ranked by a panel of three certified Arabic-writing FDEs. Each FDE assigned a numerical rank from 1 to 4 to the elements of Line Quality, Form and Proportion in the simulations. The judging was carried out according to standard forensic document examination methods, involving many painstakingly small measurements as well as more subjective observations. Fig. 1 shows the two target signatures with examples of some of the features examined by the FDEs.

The definitions of the four ranks are as follows.

Rank 1: Extremely different from the same element in the target signatures, suggesting strongly that the two signature sets were written by different people.
Rank 2: Considerably different simulated element from the same element in the target signatures, as might be likely if the two signature sets were written by different people.
Rank 3: Considerably similar to the same element in the target signatures, as might be likely if the two signature sets were written by the same person.
Rank 4: Identical or almost identical to the same element in the target signatures, suggesting strongly that the two signature sets were written by the same person.

Examples of some of the simulated signatures are shown in Fig. 2.

In actual fieldwork, of course, FDEs would seek more material than two sets of signatures. The judging in the experiment assumed that the two sets were representative of larger samples of material from the same sources. After each judge assigned a rank to each element in an individual's simulations of the signatures, the mean of the three judges' ranks was assigned to that element for that participant.

After the ranks were assigned, the data were analysed using SPSS 13.00 statistical analysis software program. Non-parametric statistical tests were used the Kruskal–Wallis test for comparing scores between more than 2 sub-groups of the three main simulated elements (Line Quality, Form, and Proportion) and the Mann–Whitney test for comparing scores of 2 sub-groups; Line Quality with Form, or Line Quality with Proportion, or Form with Proportion. The significances of mean ranks distribution differences were evaluated using the chi-square test.

Data collection. The aim of the sample collection was to gather a representative cross-section of Arabic-speaking citizens, which included fairly typical proportions of demographic characteristics such as gender, age, dominant writing handed, varied educational and occupational backgrounds, and varied skill at handwriting. Participants were seated comfortably in their normal writing position and asked to complete the standard survey form. The forms were attached to clipboards, providing the same writing surface to all participants. Each was given a fine-point, 0.7 mm, black or blue ballpoint pen, considered the best size to show fine details of handwriting. The participants were allowed to study the target signatures for about half an hour. They were not invited to make a more extended study for logistical reasons, the most important being to avoid the form being completed in other environments and on other writing surfaces.

The survey form was lined so as to provide horizontal baselines, against which alignment and slant could be measured.

3. Results and discussion

Table 1 shows the sums of the 901 mean ranks for Form, Proportion and Line Quality, assessed by a Kruskal–Wallis test. The Kruskal–Wallis test is the non-parametric equivalent of the ANOVA, which is used to compare three or more groups of sample data. It is a distribution free test in that it makes no assumption about the data being normally distributed. In this case the sample groups are the Form, Proportion and Line Quality of the simulated signatures. The results of the Kruskal–Wallis test are summarised in Table 1. There are marked differences in the rankings for Form, which has the highest ranking, with Proportion second, and Line Quality the worst. The sums of mean ranks for Form and Proportion are closer to each other than either is to the sum of mean rank for Line Quality. The Kruskal–Wallis χ^2 test shows that these differences are significant beyond the 0.001 level.

Thus the Kruskal–Wallis test shows that the relative ease with which the elements of the signatures are simulated are Form > Proportion > Line Quality, and that there are significant differences in the ability with which the participants were able to simulate these. The score in the mean rankings column indicates the success with which each of the elements of the signature was simulated, the higher the score the better the simulation. It is still necessary to locate where the

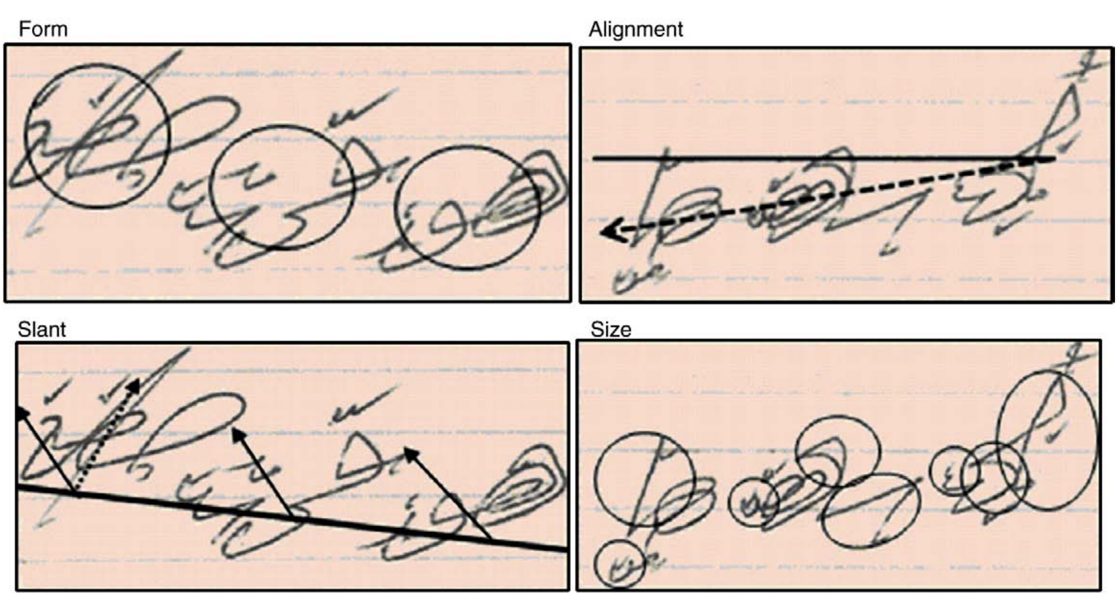

Fig. 1. The target signatures. Highlighted are examples of some of the areas examined for quality of simulation.

Please cite this article as: A. Al-Musa Alkahtani, A.W.G. Platt, A statistical study of the relative difficulty of freehand simulation of form, proportion, and line quality in Arabic signatures, Sci. Justice (2009), doi:10.1016/j.scijus.2009.08.003

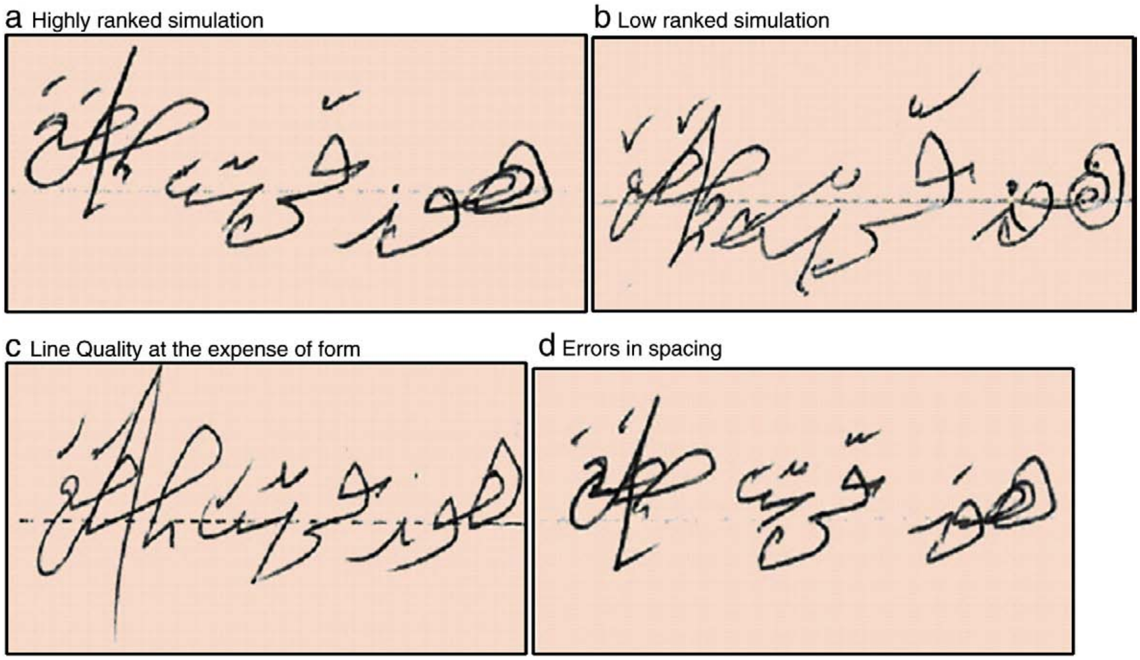

Fig. 2. Examples of some of the simulated signatures with comments on their quality.

differences in rankings are, i.e. which of the groups are significantly different from each other. To locate these differences rankings of pairs of the variables were analysed by Mann–Whitney *U* tests. The Mann–Whitney *U* test is the non-parametric equivalent of a paired *t*-test which analyses for the significance of differences in the median values of ranked data. Table 2 shows results of tests on the difference in rankings of pairs of variables. For all the pairs the differences are significant beyond the 0.001 level.

Whilst the differences are significant the data can further be analysed to assess the effect size, i.e. whether the difference detected is a large, medium or small effect [6]. This is evaluated by the *r*-value, calculated as $r=|Z/\sqrt{N}|$ where N is the number of participants, 901 in this case. Values of r above 0.5 are indicative of a large effect, if $0.5>r>0.1$ the effect is termed a medium one, and $r<0.1$ represents a small effect. Thus in Table 2 we can see that the difference between Form and Line Quality is in the region of a large effect whilst the difference between Proportion and Line Quality is a medium sized effect and that between Form and Proportion is tending towards a small effect.

These experimental results from an Arabic-writing population support the widely held view of Roman script FDEs that Form is more easily simulated than Proportion, and that both Form and Proportion are more easily simulated than Line Quality. The usual explanation given for this pattern in Roman script forgery seems to apply equally well to Arabic signatures: the larger and more consciously learned parts of writing are more readily noticed. Thus, most aspects of Form, and some aspects of Proportion, are consciously taught and learned in school, and Form, especially, tends to involve large patterns. Line Quality, by contrast, develops mainly unconsciously as writing becomes more habitual, and many differences in Line Quality are small and subtle. The well-known importance of Line Quality as a marker of forgery thus seems to be validated in a script other than Roman. An additional reason for the poor Line Quality may be the conditions of the experiment; freehand simulation with the target in view encourages copying, which usually causes poor Line Quality. Leung et al. [17] show that a Chinese sample noticed and simulated Form better than Proportion, and Proportion better than Line Quality, in a simulation task involving abstract symbols. This strikingly suggests that ranking of the three elements may transcend specific languages and scripts.

Table 1
Kruskal–Wallis test: line quality, form, and proportion.

Variable	Mean rank
Form	1554.38
Proportion	1426.30
Line quality	1075.32

$N=901$ for each variable.

4. Conclusion

In summary, this experimental study of simulation of Arabic signatures strongly supports the widespread view of expert FDEs of the relative ease of simulation of elements in Roman script. These results suggest that these effects of the three elements on ease of simulation are probably characteristic of all cursive scripts. Given the findings of Leung et al. [17], it seems likely, in fact, that they are also characteristic of non-cursive scripts, as well as symbol sets that are not functional scripts at all. The pattern seems to be extremely robust in simulation of characters and designs of all kinds.

Acknowledgements

We are grateful to the following FDEs for their time and effort in analysing the simulated signatures: Ahmad Assiry, Jubran Qushaish and Bandar Al-Hokbani.

One of us, AA, is grateful to the Counterfeiting and Forgery Department, Public Security, Ministry of Interior, Saudi Arabia for financial support.

Table 2
Mann–Whitney test showing mean rankings for pairs of variables.

	Mean rank	Z	p	r
Form	1057.61	−14.049	0.000	0.47
Line quality	745.39			
Form	947.78	−3.879	0.000	0.13
Proportion	855.22			
Proportion	1022.07	−10.244	0.000	0.34
Line quality	780.93			

Please cite this article as: A. Al-Musa Alkahtani, A.W.G. Platt, A statistical study of the relative difficulty of freehand simulation of form, proportion, and line quality in Arabic signatures, Sci. Justice (2009), doi:10.1016/j.scijus.2009.08.003

References

[1] J. Behrendt, Alzheimer's disease and its effect on handwriting, Journal of Forensic Sciences 29 (1984) 87–91.
[2] D.A. Black, Forged signatures more skillfully written than the true signatures, Journal of Criminal Law, Criminology and Police Science 53 (1) (1963).
[3] M. Boisseau, G. Chamberland, S. Gauthier, Handwriting analysis of several extra-pyramidal disorders, Canadian Society of Forensic Science 20 (1987) 139.
[4] R.A. Cabanne, Clifford Irving hoax of the Howard Hughes autobiography, Journal of Forensic Sciences. 20 (1975) 05–17.
[5] B.B. Carney, A new tremor in handwriting (Book Review), International Journal of Forensic Document Examiners 1 (1995) 75–77.
[6] J. Cohen, Statistical Power Analysis for the Behavioural Sciences2nd Ed, Lawrence Erlbaum Associates, Inc, Hillsdale NJ, 1988.
[7] P. Collins, I. Lee, P. Tyrer, Finger tremor and extrapyramidal side effects of neuroleptic drugs, The British Journal of Psychiatry 134 (1979) 488–493.
[8] J.V.P. Conway, Personal and personality in anonymous letters, Journal of Forensic Sciences 14 (1959) 18–33.
[9] G. Epstein, Examination of the Josef Mengele handwriting, Journal of Forensic Sciences 32 (1987) 100–109.
[10] J.J. Harris, How much people write alike: a study of signatures, Journal of Criminal Law, Criminology and Police Science 48 (6) (1958) 647–651.
[11] W.R. Harrison, Suspect Documents their Scientific Examination, Nelson-Hall Inc, 1981.
[12] A. Herkt, Signature disguise or signature forgery, Journal of the Forensic Sciences Society 26 (4) (1986) 257–266.
[13] O. Hilton, Scientific Examination of Questioned Documents, revised ed. Elsevier, New York, 1993.
[14] O. Hilton, Influence of the Age and Illness on Handwriting Identification Problems, Paper Presented on the Annual Meeting of the American Society of Questioned Document Examiners. Colorado. USA, 1975.
[15] J.J. Horan, The Size of Forged Signatures, Presented at the Meeting of the American Society of Questioned Document Examiners. Chicago, 1985, 1985.
[16] R.A. Huber, A.M. Headrick, Handwriting Identification: Facts and Fundamentals, CRC press LLC., USA, 1999.
[17] S.C. Leung, Y.S. Cheng, H.T. Fung, N. Poon, Forgery 1 — simulation, Journal of Forensic Sciences 38 (2) (1993) 402–412.
[18] J.F. Masson, An Evaluation of Simulated and Traced Signatures and Consideration of the Potential for Determination of the Writer's Identity, Presented at the Meeting of the American Academy of Forensic Sciences, Nashville, 1996.
[19] J.F. McCarthy, Two Penmen's Ability in Simulation Signatures, Presented at the Meeting of the International Association of Forensic Sciences, Oxford, 1984.
[20] J. Nickell, The Jack the ripper diary: history or hoax? International Journal of Forensic Document Examiners 3 (1997) 59–63.
[21] A.S. Osborn, Questioned Documents, Second EditionPatterson Smith, New Jersey, 1929.
[22] D.K.S. Puri, Study of a copied forgery, Journal of Security Administration. 3 (1980) 79–87.
[23] K. Rendell, Forging History, University of Oklahoma Press, USA, 1994.
[24] R.M. Ruff, S.B. Parker, Gender and age specific change in motor speed and eye hand coordination in adults: normative values for the finger tapping and grooved pegboard tests, Perceptual and Motor Skills 76 (1993) 1219–1230.
[25] Robert Saudek, Experiments with Handwriting, George Allen & Unwin, London, 1928.
[26] R.N. Totty, Skilled Copies of Signatures, Presented at the Meeting of the American Society of Questioned Document Examiners, Chicago, USA, 1995.

Please cite this article as: A. Al-Musa Alkahtani, A.W.G. Platt, A statistical study of the relative difficulty of freehand simulation of form, proportion, and line quality in Arabic signatures, Sci. Justice (2009), doi:10.1016/j.scijus.2009.08.003