Journal of the American Society of Questioned Document Examiners, Inc.
Volume 17, Number 1, June 2014

Arabic-Illiterate Forensic Handwriting Analysis: A Pilot Study to Further Investigate the Ability of Arabic-Illiterate Examiners to Judge the Accuracy of Simulations of Arabic Signatures, Compared with Arabic-Literate Examiners

Abdulaziz Al-Musa Alkahtani, MSc, PhD, MFSSoc

Questioned Document Department, Public Security Saudi Arabia, E-mail: almusa@live.com

This article extends the study in Al-Musa (2010) examining whether forensic handwriting examiners (FHEs) who are illiterate in Arabic can detect Arabic forgeries as accurately as FHEs who have native fluency in Arabic. Because large numbers of real forgeries are difficult to find, the forgery data in these two pilot studies consisted of simulations, produced by 100 native Arabic writers, of two Arabic signatures. Two experienced Arabic-literate FHEs and two experienced Arabic-illiterate FHEs then judged the accuracy of these simulations, using standard forensic handwriting analysis methods. Their judgments of four narrow elements of the simulations (Size, Spacing, Arrangement, and Slant), as well as two of the broader elements studied in the 2010 article (Form and Line Quality), were compared. As in the earlier study, the judgments of Arabic-literate FHEs did not differ significantly from those of Arabic-illiterate FHEs ($p = 0.05$). Thus, a second statistical study adds support to a widespread view held by FHEs, based on anecdote and experience, that literacy in a script is not needed to detect forgery in that script. Other considerations, however, suggest that a native writer of a script may have advantages over an illiterate in that script in conducting forensic handwriting analysis.

This pilot study follows up questions raised in the earlier investigation by Al-Musa (2010) of the ability of forensic handwriting examiners (FHEs) to judge the accuracy of simulations written in scripts with which they were not familiar. Prior to the earlier article, anecdotal evidence had accumulated suggesting that forensic handwriting examination techniques could be used to identify forgeries in scripts unknown to the FHEs, as well as in abstract designs, drawings, and paintings (see discussion in Al-Musa 2010 and also Hensel, Khan, & Dizon 1973; Hanna 1989; Leung, Cheng, Fung, & Poor 1993; Miller 1995; Huber & Headrick 1999). This had become a commonly held assumption in the field of handwriting analysis, and the study in Al-Musa (2010) was undertaken in the hope of either disproving it or placing it on a firmer foundation, in part to provide justification for it in courts of law.

In Al-Musa (2010), a sample of 100 native Arabic-speakers and Arabic-writers attempted to simulate two Arabic signatures under highly similar conditions. It is extremely difficult for students of forgery to find a large sample of actual forgeries, and therefore simulation, in which a subject attempts to imitate the appearance of handwriting without the anxiety and time pressure of real forgery, was used as the next best approximation. After the simulations were produced, four FHEs, two Arabic-literate (native speakers and writers), and two Arabic-illiterate (native English speakers and writers), compared each attempt with the model signatures, and ranked it for accuracy of three elements: Form, Proportion, and Line Qual-

ity. The study found that the differences in rankings between the Arabic-literate and Arabic-illiterate FHEs did not differ statistically at the level of $p = 0.05$. Thus, it appeared probable that forensic handwriting analysis techniques could be applied as well by FHEs who were literate as by FHEs who were illiterate in at least some scripts, as had been widely believed by FHEs.

It was suggested in the 2010 article that research on a wider range of elements, languages, and scripts should be carried out, to further test and possibly generalize these results. In the present pilot study, the broadest of the three elements, Proportion, was broken into four narrower sub-elements, on the hypothesis that certain sub-elements might be more difficult than others for Arabic-illiterate FHEs to judge, and that this difference might have been lost in the overall rankings of Proportion. The sub-elements were Size, Spacing, Arrangement, and Slant.

The four elements are explained below, using the Arabic signatures that were simulated in both studies. The first element is Arrangement, which in this study was mainly concerned with the alignment of the baseline of the writing with horizontal lines on the survey sheet, although it also took into account gross errors of the general placement of the writing in the space provided. Figure 1 shows the model signatures against the lined background on which they were presented. The red lines in the figure are aligned with the horizontal lines, while the green lines indicate the degrees to which the baselines of the two signatures are out of alignment with the horizontal. As can be seen, the baseline of Signature 1 rises upward from the right (where Arabic writing starts), while the baseline of Signature 2 falls downward from the right. In this study, the FHEs judged the accuracy of the degrees of alignment of the writing baselines with the true horizontal lines in the simulations.

Arrangement seemed an especially good candidate for this investigation because the author has noticed tendencies for the baseline of a line of Arabic handwriting (which is written from right to left) to gradually descend toward the left, while Roman script such as English handwriting (written from left-to-right) often seems to rise forward toward the right. The expectations of English and Arabic FHEs with respect to the arrangement of the baseline of writing might therefore differ and affect their perceptions of the accuracy of simulations.

The second element, Slant, refers to the slants of parts of letters or whole letters relative to the baselines of the writing. Letters may tend to slant forward or backward in the writing of different individuals, and particular letters may be slanted differently as a rule, or in different environments, within the same individual's handwriting. In Figure 2, the usual slant of most of the letters in the two signatures, indicated by the red lines, is forward (toward the left), relative to the baselines of the signatures, shown by the yellow lines. However, the vertical letter at the beginning (far right) of Signature 2 and the staffs of other letters slant backward (to the right), as indicated by the green arrows, relative to the slants of the writing baselines. In this pilot study, the FHEs judged the accuracy of the degrees of slants of letters and parts of letters, relative to the baselines of the signatures in the simulations.

Like Arrangement, Slant seemed an element that might cause differences in judgments of accuracy for English-writing and Arabic-writing FHEs. The author has noticed tendencies for writers of Arabic (which is written from right to left) to slant letters forward, toward the left, and for writers of English and European (Roman script) languages (written from left to right) to slant letters forward toward the right. This might affect the expectations and perceptions of English-writing and Arabic-writing FHEs.

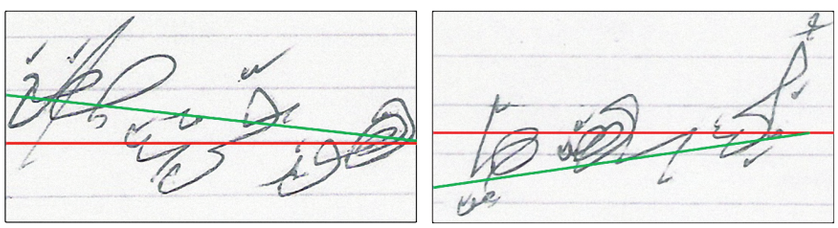

Figure 1: The two model signatures with horizontal baselines indicated by the red lines and the alignments of the baselines of the signatures indicated by the green lines.

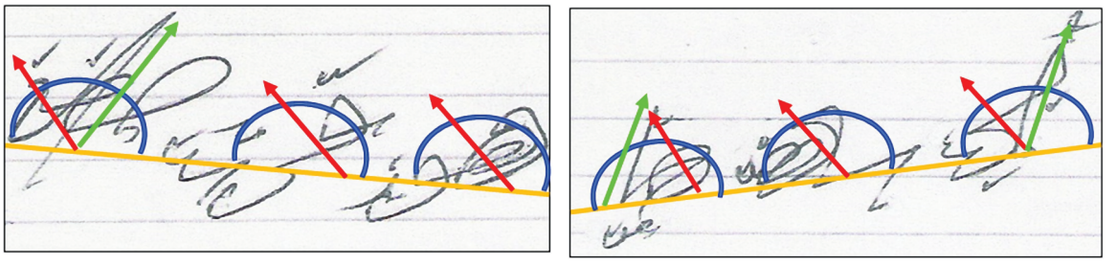

Figure 2: Forward slant of some letters (red lines) and backward slant of others (green arrows), relative to the baselines of the writing (yellow lines).

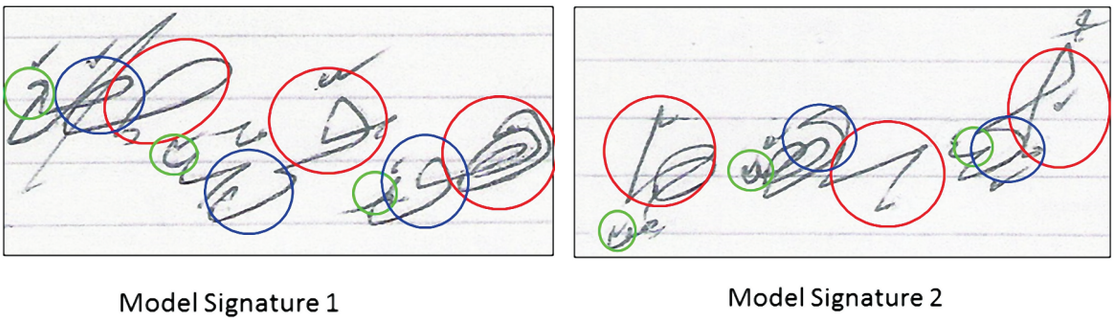

Figure 3: The two model signatures with especially large letters circled in red, medium-sized letters circled in blue, and especially small letters circled in green.

The third element, Size, will be readily understood by readers. Individuals writing Arabic tend to produce letters and words of different sizes, especially in different environments, as do writers of Roman script. Figure 3 shows especially large letters circled in red, medium-sized letters circled in blue and especially small letters circled in green. As can be seen, in each writing form (a line of writing in which the cursive line breaks before and after, roughly corresponding to a word), the initial letters (on the right side) are especially large, while the final letters (left side) are especially small.

Figures 4 and 5 illustrate the element of Spacing, both between words and between letters. This element receives relatively little attention in forensic analysis of English, but the author has found it to be more significant in some forensic analyses of Arabic. In Figure 4, the green triangles in Model Signature 2 indicate the spaces between the words in that signature; the bases of the triangles show the spaces between the words. The same triangles (in red) are superimposed on the spaces between words in Model Signature 1, showing that the spaces between the same words are much smaller than in Signature 2. In fact, the ends of the words in Signature 1 actually overlap with the beginnings of the next words, whereas real spaces exist between the words in Signature 2. The FHEs in the study judged the extent to which the simulators captured that difference.

Figure 5 shows spacing between letters. Again, the cursive writing line breaks between each word. Each letter within each word is circled. It can be seen that there are spaces between letters in the three words in Signature 1, but that letters within the three words in Signature 2 usually overlap. Signature 1 has rather loose, widely spaced letters with long connecting lines between them, while Signature 2 has rather narrow, cramped letters with very short connecting lines between them.

The design of this study closely resembles that in Al-Musa (2010). Six simulations of each of the two signatures were elicited from each of 100 Arabic-literate writers and speakers. Four FHEs, two Arabic-literate and two Arabic-illiterate, then ranked the accuracy of the simulations of the four target elements. In order to compare the Arabic-literate and Arabic-illiterate FHEs, mean ranks (Mann-Whitney U) were then calculated

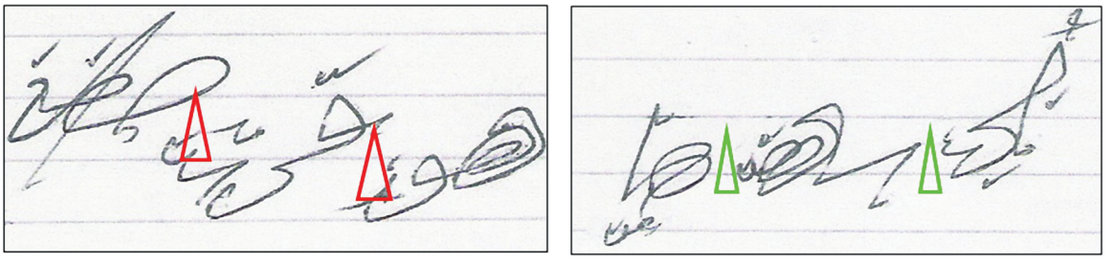

Figure 4: Spacing between words: longer spaces between words in Signature 2 and negative spaces between words in Signature 1.

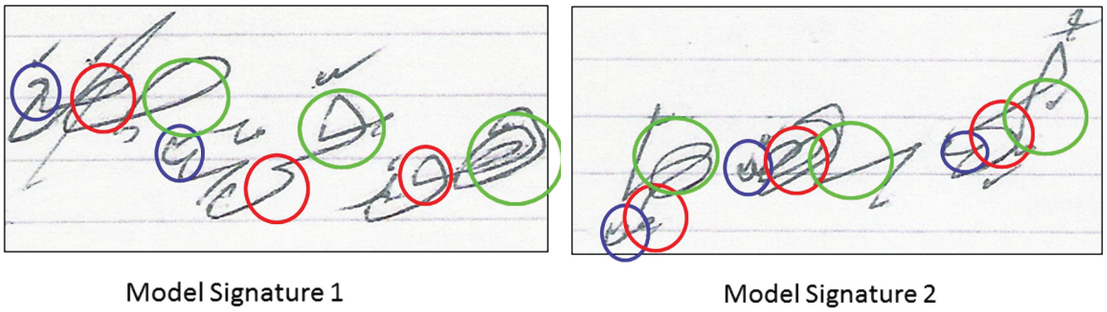

Figure 5: Spacing between letters: The boundaries between letters are indicated by circles. Within words, letters are widely spaced in Signature 1, but usually overlap in Signature 2. Different colors help to distinguish circles.

separately for the Arabic-literate FHEs and the Arabic-illiterate FHEs for each of the four elements, using the Mann-Whitney U Test in the SPSS 13.00 software package.

Methods and Materials

This pilot study was carried out with the same sample of 100 Arabic-literate simulators and the same four FHEs (two Arabic-literate and two Arabic-illiterate) who participated in the first study (Al-Musa 2010). The FHEs were all certified and had had at least four years of experience as professional forensic handwriting analysts. The same sample of simulations used in the 2010 Al-Musa study also supplied the data on these four Proportional elements. The simulation data had been collected by the author in Saudi Arabia, Egypt, and Morocco, mostly from groups provided by businesses, colleges, offices, and other institutions. The subjects made their simulations under very similar physical conditions, using the same type of fine-point, 0.7 mm black or blue ballpoint pen, and filling out identical survey instruments, fastened to clipboards that provided the same type of surface to each subject. Subjects were given about 20 minutes to study the model signatures, and about 10 minutes to make three freehand simulations of each of the two model signatures.

The finished survey instruments were collected and circulated to the four FHEs, so that each could rank the accuracy of the simulations of each subject for each element, using standard forensic handwriting analysis techniques. They were allowed as much time as necessary to complete this task, and the Arabic-illiterate FHEs took considerably longer than the Arabic-literate FHEs, in some cases months longer, as in the earlier analysis of Al-Musa (2010). The survey instrument, shown in Figure 6, contained spaces to collect data for several research projects. The most relevant spaces for this study were (1) the two model signatures at the left near the top and the boxes below them, where the simulations were written and (2) the rows at the bottom where FHEs entered the ranks of simulation accuracy for different elements. The boxes on each

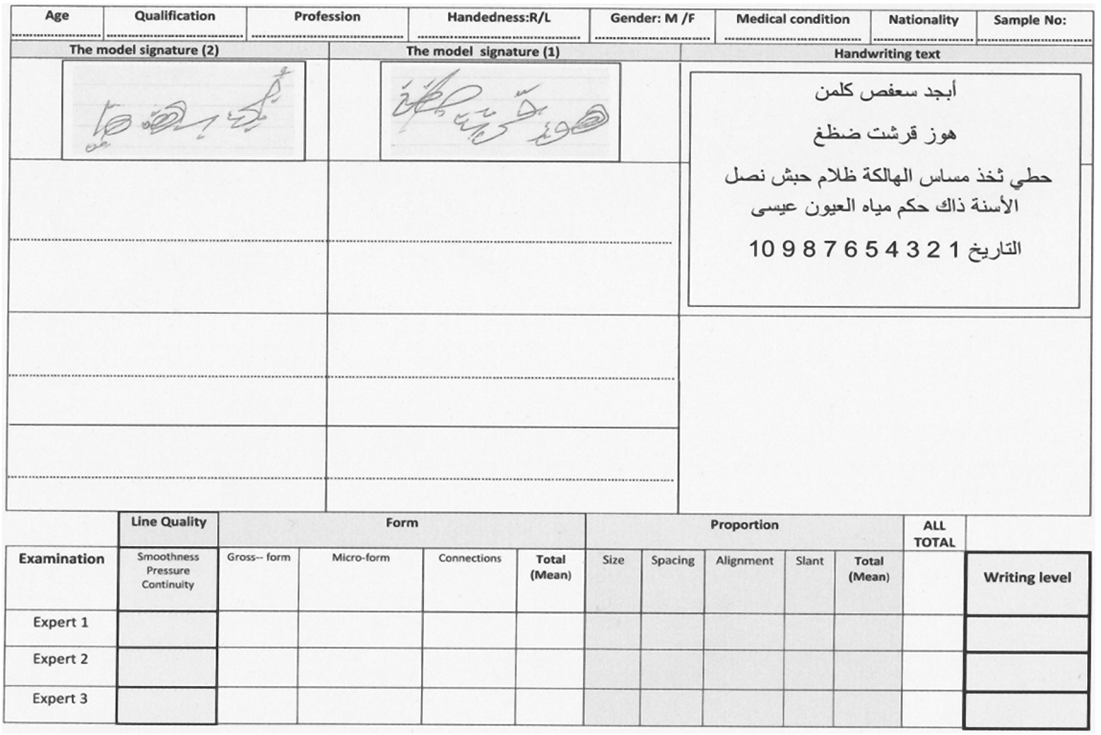

Figure 6: The Survey Instrument.

row in which ranks for the four elements were entered are the shaded columns toward the right. The FHEs assigned a simulation-accuracy rank to each of the four elements for each subject (that is, for each survey instrument).

The ranks used by the FHEs were defined as follows:

- Rank 1: Extremely different from the element in the model signature;
- Rank 2: Quite different from the element in the model signature;
- Rank 3: Similar to the element in the model signature;
- Rank 4: Identical or almost identical to the element in the model signature.

Results

In order to compare the rankings of the two sets of FHEs for simulations accuracy of the four elements (Size, Spacing, Placement, and Slant), the mean ranks for each element assigned by each set of FHEs were calculated. The significances of the differences between the two sets of FHEs were then determined, using the Mann-Whitney U test with a significance level of 0.05 (SPSS 13.00). Data from the first study on the elements of Form and Line Quality (Al-Musa 2010), which were collected from the same simulators and ranked by the same four FHEs, were added to the analysis. Table 1 shows the significance of the differences between rankings of the two sets of FHEs for each of the six elements. As can be seen, none of the asymptotic 2-tailed p values were less than 0.05, indicating that the Arabic-literate and Arabic-illiterate accuracy ratings were not significantly different for any of the elements.

It is interesting that the largest difference between the two sets of FHEs, as indicated by the p value, occurred in their rankings of Slant, and the second largest in their rankings of Arrangement. It was noted in the explanations of Slant and Arrangement in the first section of this article that both elements might be especially vulnerable to errors made by FHEs who were illiterate in the languages they were examining. FHEs who work with English are accustomed to a left-to-right script that tends to have letters that slant toward the right, and that tends to rise as the whole body of writing moves forward toward the right. FHEs who work with Arabic, a right-to-left script, are used to letters that slant toward the left, and lines of writing that tend to descend toward the

Element	Line Quality	Form	Size	Spacing	Arrangement	Slant	Total quality of freehand simulation
Mann-Whitney U	4933.000	4709.500	4785.500	4720.500	4890.000	4487.500	4765.500
Wilcoxon W	9983.000	9759.500	9835.500	9770.500	9940.000	9537.500	9815.500
Z	-.174	-.755	-.572	-.733	-.280	-1.296	-.575
Asymptotic Significance (2-tailed)	.862	.450	.568	.463	.779	.195	.565

Table 1: p values for the differences in ranks assigned to the two sets of FDEs for each of the six elements.

left. This suggests only that bias based on experience and expectations might be more likely to introduce error into judgments of the accuracy of simulations of Slant and Arrangement when FHEs are used to languages written in opposite directions from their own. There is no evidence of significant differences in these data, of course.

Table 2 shows the Mann-Whitney U mean ranks and the sums of ranks for each FHE-element category. As scrutiny of the table reveals, the mean ranks assigned by the Arabic-writing FHEs were consistently slightly higher (judging slightly greater accuracy) than those assigned by the Arabic-illiterate FHEs. This might reflect a slight overcorrection by the Arabic-illiterate for their lack of knowledge of Arabic, in finding more, or more serious, errors. But, again, the difference is not significant.

Conclusions

The findings in Al-Musa (2010) and in this pilot study seem to imply that a trained, experienced FHE is likely to be able to identify forgeries in an unfamiliar script (at least if it is cursive) about as well as a similarly qualified FHE who has native familiarity and fluency in the script. However, common sense suggests some caveats and more related studies are needed on wider range of elements, languages, and scripts to further test and possibly generalize these results, and to establish whether the technique and study protocol are effective and realistic.

	FHE	N	Mean Rank
Line Quality	Arabic FHE Non-Arabic FHE Total	100 100 200	101.17 99.83
Form	Arabic FHE Non-Arabic FHE Total	100 100 200	103.41 97.60
Size	Arabic FHE Non-Arabic FHE Total	100 100 200	102.65 98.36
Spacing	Arabic FHE Non-Arabic FHE Total	100 100 200	103.30 97.71
Arrangement	Arabic FHE Non-Arabic FHE Total	100 100 200	101.60 99.40
Slant	Arabic FHE Non-Arabic FHE Total	100 100 200	105.63 95.38
Total quality of freehand simulation	Arabic FHE Non-Arabic FHE Total	100 100 200	102.85 98.16

Table 2: Mann-Whitney U mean ranks and sums of ranks for FHE-element categories.

In the first place, the FHEs in both studies took as long to make their judgments as they needed, and the Arabic-illiterate FHEs took several months longer to get their results back to the investigator than the Arabic-literate FHEs. It is certainly plausible that it would take them longer to analyze the differences between model signatures and simulations, and such a delay might be inconvenient or worse in a real criminal case. In the second place, the scale of simulation-accuracy ranks used in both studies offers only four categories; a larger number of categories might have created stronger differences between the two sets of FHEs. In the third place, these tests involved analysis of only the physical aspects of the script, which is a vital part of forensic handwriting analysis, of course, but not the only part. Understanding facts about the use of handwriting in unfamiliar cultures may also be important in reaching judgments about forgery. (An example would be knowledge about the common custom in Arab culture, with its long interest in calligraphy, of inventing and using abstract designs for signatures in some situations. A native Arabic writer might have a better idea of when such a signature would or would not be appropriate.) Thus, further research should include studying the actual performances of literate and illiterate FHEs in real cases of suspected forgery. These would necessarily be case studies, anecdotal rather than statistical, but might give rise to insights that could be tested using correlational or experimental methods. The aim would be to discover situations and combinations of cultures in which FHEs from foreign cultures might have special difficulties analyzing suspected forgeries. Although this pilot study was undertaken partly to apply correlational statistics to material that had previously been demonstrated anecdotally, it is useful to remember that there is room for descriptive science, as well as correlational and experimental science, in the accumulation of knowledge, as has always been the case in all scientific fields. It would also be valuable to extend this type of correlational study, elucidated with case studies, to scripts that have even fewer traits in common, such as comparisons of FHEs who are native writers of cursive Arabic or Roman scripts with printed Oriental scripts.

References

Al-Musa Alkahtani, A. (2010). The Ability of Forensic handwriting Examiners to Judge the Quality of Signature Simulations in an Unfamiliar Writing System. Journal of the American Society of Questioned Document Examiners. Journal of the American Society of Questioned Document Examiners 13 (2).

Al-Musa Alkahtani, A. and Platt, A. (2009). Relative Difficulty of Freehand Simulation of Four Proportional Elements in Arabic Signatures. Journal of the American Society of Questioned Document Examiners Volume 12, Number 2.

Daubert v. Merrell Dow Pharmaceuticals (1993) 509 U.S. 579.

Hanna, G. A. (1989). "A Preliminary Classification of the Writing Elements of Chinese Characters." Journal of Forensic Sciences 34 (2): 439-448.

Harris JJ. How much do people write alike?. (1958). Journal of Criminal Law. Criminology and Police Sciences ;48:6.

Hensel, E.R., Khan, I.A., and Dizon, J.F. (1973). "Forensic examination of peculiar writing systems." Journal of Forensic Sci. Society 13: 143-152.

Hilton O. Scientific examination of questioned documents. (1993). revised ed. New York: Elsevier North Holland Inc.

Huber, R.A., and Headrick, A.M. (1999). "The Scope of Document Examination." In Handwriting Identification: Facts and fundamentals, 327-330. Boca Raton, London, New York: CRCC Press LLC.

Leung, S.C., Cheng, Y.S., Fung, H.T., and Poor, N. (1993). "Forgery I – simulation." Journal of Forensic Sciences 38 (2): 402-412.

Miller, L.S. (1995). "Identification of human figure drawings through questioned document examination techniques." Forensic Science International 72: 91-105.