Journal of the American Society of Questioned Document Examiners, Inc.
Volume 14, Number 2, December 2011

Skill Characteristics of Forensic Handwriting Examiners Associated with Simulated Handwritten Text

Carolyne Bird[1, 2], Reinoud D. Stoel[3], Bryan Found[1,4] and Douglas Rogers[1]

[1]Handwriting Analysis and Research Laboratory, School of Human Biosciences, La Trobe University, Bundoora VIC 3086, AUSTRALIA

[2]Document Examination Section, Forensic Science SA, Adelaide SA 5000, AUSTRALIA

[3]Netherlands Forensic Institute, The Hague, The Netherlands

[4]Document Examination Team, Victoria Police Forensic Services Department, Macleod VIC 3085, AUSTRALIA

The assessment of the process of production of handwriting (naturally written, disguised or simulated) is an important step in forensic examinations and may impact on any authorship opinion offered. However, there is currently little empirical data on the skill of forensic handwriting examiners in discriminating between disguised and simulated writing processes. The results reported here form part of a larger investigation to that end. The trial consisted of 100 pairs of handwriting samples, each with a naturally written comparison sample, and a questioned sample that was either disguised by the comparison writer, or written by another writer attempting to simulate the comparison writer's handwriting features. The simulated writings were made with or without practice and with or without a direct model of the target text in the comparison writer's handwriting. Analysis revealed no significant differences between the correct or misleading (erroneous) scores of any of the combinations of simulated writing types. Simulated writings attracted high misleading and inconclusive scores indicating that the predictor features that FHEs use to form opinions on the un-natural writing samples in this study were not effective.

Introduction

In the past fifteen years the issue of validation of forensic handwriting examiners' (FHEs) skills has been investigated (Kam, et al. 1994, 1997, 2001, 2010; Found, et al 1999; Sita, et al. 2002; Dyer, et al. 2006). Controlled blind studies have found that FHEs do have expertise in relation to opinions on questioned handwriting and signatures. More recent research on disguised and simulated writings has identified these as problem writing types for FHEs (Found and Rogers, 2005, 2008). It has been reported that un-natural writing types attract higher incorrect and inconclusive scores on authorship opinions than genuine writing does. If FHEs can identify a sample of questioned writing as the product of an un-natural (disguised or simulated) writing process, they can express this as a ''first-stage'' opinion. Without empirical evidence as to the validity of proceeding on to express an opinion as to authorship (a ''second-stage'' opinion) the probative value of opinions of this type remains unknown. Previous research reported by the authors found that FHEs have a skill over laypeople in distinguishing between natural and disguised handwriting behaviours (Bird, et al. 2010a). Like their expertise for authorship opinions, this expertise is characterised by FHEs conservatism, leading to a smaller percentage of misleading responses when calling writing as either naturally written or disguised, when compared to laypersons. However, there is no published empirical data on the skill of FHEs in discriminating between simulated and disguised handwritten text. The results reported here form part of a larger study to that end, directly addressing the impact of practice and the use of a direct model for simulation on forensic handwriting experts' ability to detect simulated handwriting. The purpose of this study is to further investigate examiners skill at assessing the process of production of un-natural writings and focuses on any effects that might arise when FHEs are attempting to discriminate between disguise and simulation behaviours resulting from the strategy used by simulators in copying writing samples.

Methods and Materials

Participants

The results reported here are from FHEs who provided independent opinions on the process of production of supplied handwriting samples. Thirty-five FHEs from fourteen countries took part in the trial (seven of these countries do not have English as an official language). All of the participants have been authorised by their employers to release opinions regarding the authorship of questioned handwriting and signatures. At the time of data collection, thirty FHEs were working in government laboratories and five were working privately.

Materials

The trial consisted of 100 pairs of handwritten text, each with a naturally written comparison sample (100 different writers), and a questioned sample that was either disguised by the comparison writer, or written by another writer attempting to simulate the comparison writer's handwriting features. Writings were made using one of two makes of ballpoint pen and the same make of white paper. The samples were scanned at 600dpi and laser-printed into a booklet as well as converted into high-resolution PDF files.

Disguised writings: No directions were given to the handwriting providers as to how to disguise their handwriting.

Simulated writings: Fifty-nine writers provided the sixty-six simulated samples in the trial. Simulators were provided with target writing of the comparison writer that was either the same text to be simulated (direct model), or collected/course of business writings from which they had to create the standard text. Simulated samples based on these two groups of target writings were either made with, or without practice. A period of one week was allowed for practice. No conditions were set for the number of attempts to be made prior to providing the final sample. Tracing was not permitted. Although the models provided to simulators contained the exact text included in the trial, the models were not used as comparison samples in the trial.

Procedure

Participants were provided with a sample booklet of laser-printed images of the handwriting samples and a DVD containing PDF files of the samples. Each participant was given an answer-recording booklet and informed that each sample pair consisted of a naturally written comparison sample and a questioned sample that was either disguised by the comparison writer, or simulated by a different writer attempting to copy the handwriting features of the comparison writer. Participants were asked to make an assessment as to whether the questioned sample was disguised, or simulated, by entering the appropriate code (D, S or I (for an inconclusive opinion)) in the answer booklet. If an inconclusive opinion was given, the participant was also required to record a forced opinion (D or S) for the handwriting sample; however, not all participants did so. Only the unforced opinions are included in this paper. In addition to an opinion on the un-natural writing process, the features the participant observed that lead them to their conclusion were requested to be recorded in the free-text box corresponding to each of the questioned samples in the answer booklet. An analysis of this aspect of the trial will be reported separately.

Ethics Approval

Approval for this study was obtained from the La Trobe University Human Ethics Committee on the basis that the handwriting providers and test participants gave full consent for samples of their handwriting, or research data provided by them respectively, to be included in published material, on the condition that neither their name nor any other identifying information be used.

Analysis

Participants' responses were marked as correct, misleading or inconclusive, and analysed to produce scores. Four of the participants responded with only inconclusive opinions and have been excluded from this analysis. Since the same samples were examined by the group of experts, and since we are interested here in aspects of the handwriting itself, we aggregated over examiners[1]. To determine whether there were differ-

[1] In a future study we will investigate the data in a slightly different way that will allow us to compare the difficulty of the handwriting samples (i.e. the proportion of examiners that gave a correct score for the specific handwriting sample) to the performance of the examiners (i.e. the proportion of correct scores for each examiner over all handwriting samples) by means of a random effect analysis. In the current study a standard ANOVA was used. We refer to Tabachnick & Fidell (2007a, 2007b) for introductory descriptions of both techniques.

Simulation condition	Practice	No practice
Model	$n_{samples}$=18; n_{FHE}=31; $n_{opinions}$=387	$n_{samples}$=17; n_{FHE}=29; $n_{opinions}$=352
No Model	$n_{samples}$=13 n_{FHE}=31; $n_{opinions}$=252	$n_{samples}$=18; n_{FHE}=30; $n_{opinions}$=359

Table 1 Number of samples, FHEs and conclusive opinions expressed in the different simulation condition groups

Simulation condition	Practice	No practice	Total
Model	.557	.527	.543
No Model	.498	.634	.577
Total	.532	.582	.559

Table 2 Mean proportion correct for conclusive opinions in each of the 2x2 conditions

Simulation condition	Practice	No practice	Total
Model	.307	.332	.319
No Model	.375	.357	.364
Total	.335	.345	.340

Table 3 Mean proportion of inconclusive opinions in each of the 2x2 conditions

ences in scores between the different conditions a 2x2 analysis of variance (ANOVA; IBM SPSS Statistics version 19) was performed.

Results and discussion

There were a total of sixty-six simulated samples in the trial. These fell into four groups based on the conditions under which they were created: direct model of target handwriting with practice time, direct model of target handwriting without practice time, no direct model of target handwriting (i.e. collected samples of target handwriting) with practice time, and no direct model of target handwriting without practice time. Table 1 gives the number of samples, number of FHEs, and the number of conclusive opinions expressed in each of the simulation condition groups.

An overview of the mean proportions correctly called in each of the four conditions is given in Table 2 (calculated after having removed the inconclusive opinions). The mean proportions of inconclusive opinions are presented in Table 3.

With respect to FHE opinions on the handwriting samples, the results show a small, though significant interaction between model and practice ($F(62,1)=4,28$; $p=.043$; $\alpha=.05$). Although the differences are relatively small, the highest correct scores are obtained in the 'no practice, no model' condition, while the lowest correct scores are obtained in the 'practice, no model condition'. With respect to the proportion of inconclusive opinions the results show a small but significant main effect for model ($F(62,1)=4,97$; $p=.029$; $\alpha=.05$), with the highest proportion of inconclusive opinions obtained in the 'no model' condition.

While the differences between the conditions are relatively small, the results do clearly show FHEs' are not skilled in the task of identifying whether a questioned sample was simulated rather than disguised, as evidenced by the high percentage misleading called (in total 44%) scores. The high percentage inconclusive scores (in total 34%) show that FHEs, in this trial, were not confident in the task either.

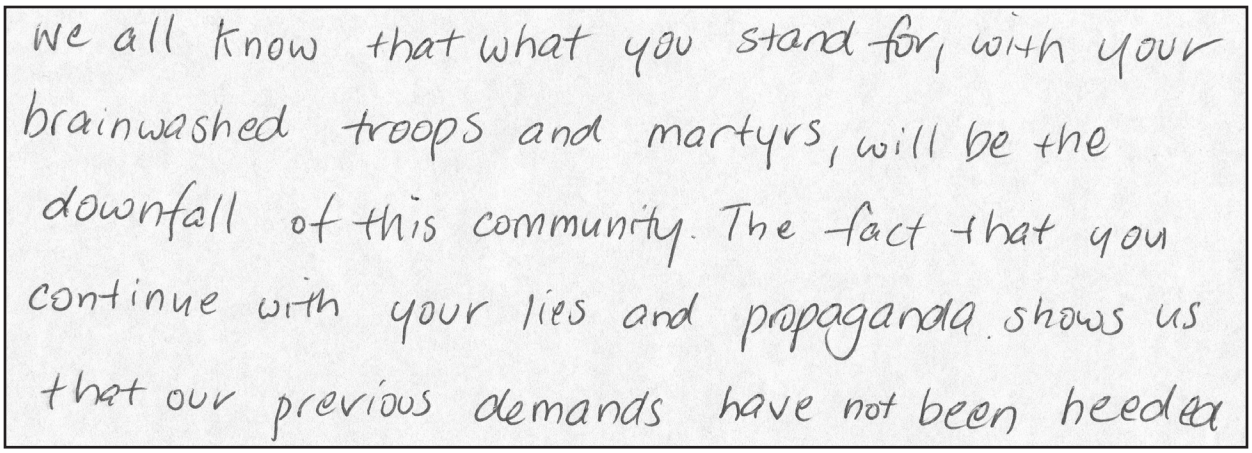

Figure 1a Naturally written comparison sample

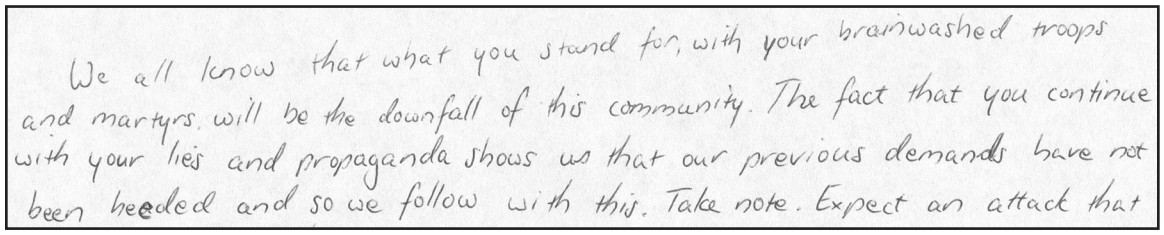

Figure 1b Questioned simulated sample (no model, practice) that attracted high inconclusive (54.8%) and misleading called (71.4%) scores

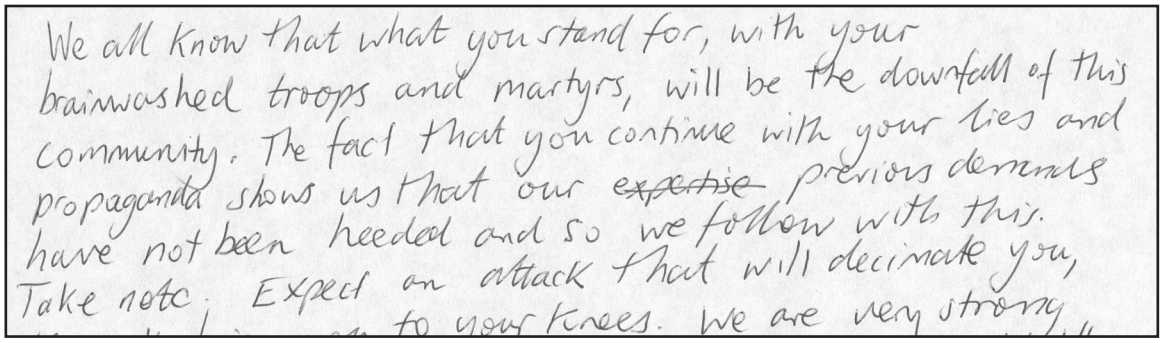

Figure 2a Naturally written comparison sample

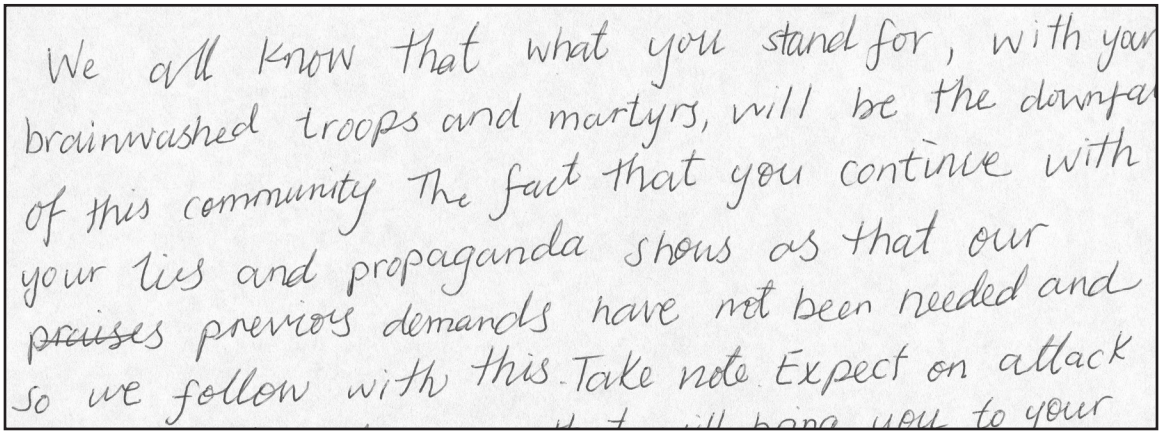

Figure 2b Questioned simulated sample (model, no practice) that attracted an average inconclusive score (32.3%) and a high correct called score (90.5%)

The task of copying the exact text of the target handwriting in the form of a direct model could be assumed to be a far more straightforward task than having to search out the individual letterforms or words from a collection of course of business writings. Simulators, having to construct text by searching for appropriate characters or words amongst collected writings, may produce samples that are different to those produced under the direct model condition. Writings of this type may provide features to FHEs that may or may not assist them in making the determination that the sample is indeed simulated (rather than disguised). However, regardless of simulation strategy, FHEs performance remained poor as characterised by the high percentage misleading called and percentage inconclusive scores. The slightly lower inconclusive scores for simulations executed using a direct model indicate that FHEs were more confident in expressing opinions on these samples. The lack of differences between the correct and misleading called scores shows that this confidence does not translate into an increased skill. On the contrary, it implies that the predictor features FHEs used to form opinions on the un-natural writing samples were not effective.

The uncertainty and error associated with simulated samples across all simulation conditions may stem from a number of sources. Poorly performed simulations that do not resemble the comparison handwriting may lead the FHE towards the opinion that the questioned sample is disguised. However, poorly and skilfully performed simulations alike are liable to exhibit subtle dissimilarities, which will tend to point towards a writer other than the comparison writer (i.e. simulation). Simulations that display both gross and subtle dissimilarities provide a mixed signal and presumably result in inconclusive opinions.

For the purpose of illustrating different profiles of FHE opinions, Figures 1 and 2 depict portions of two of the trial sample pairs. The questioned simulated sample Figure 1b, which was created without a direct model but with practice, displays both gross and subtle dissimilarities and attracted high inconclusive (54.8%) and misleading called (71.4%) scores. Figure 2b shows a simulated sample that was written using a direct model but no practice. The presence of similarities in gross features with subtle dissimilarities lead to a high correct called score (90.5%), with an average inconclusive score (32.3%).

Conclusion

The results of this study provide further evidence that un-natural writing behaviours are problematic for FHEs. Their ability to discriminate between simulated and disguised writings using traditional methodological approaches is poor as characterised by high inconclusive and misleading scores. Given the results of our prior research (Bird, et al, 2010a, 2010b; Found & Rogers, 2005), the data presented here provides further evidence that when it is opined that a questioned handwriting sample is un-naturally written, and where both disguise and simulation propositions are being considered, there is no empirical evidence that would support a second-stage authorship opinion being expressed.

Acknowledgments

We are grateful to the handwriting providers and forensic document examiners who generously gave their time to participate in the study.

References

Bird, C.; Found, B.; Rogers, D. (2010) Forensic document examiners' skill in distinguishing between natural and disguised handwriting behaviors. Journal of Forensic Sciences. Vol. 55, No. 5, pp. 1291–1295.

Bird, C.; Found, B.; Rogers, D. (2010) Forensic Handwriting Examiners' skill in detecting disguise and simulation behaviour from handwritten text samples, Presented at the 20th International Symposium on the Forensic Sciences, September 5 – 9; Sydney, Australia.

Dyer, A.G.; Found, B.; Rogers, D. (2006). Visual attention and expertise for forensic signature analysis. Journal of Forensic Sciences, Vol. 51, pp. 1397–404.

Found, B.; Sita, J.; Rogers, D. (1999). The development of a program for characterising forensic handwriting examiners' expertise: signature examination pilot study. Journal of Forensic Document Examination, Vol. 12, pp. 69–80.

Found, B., Rogers, D. (2005) Problem types of questioned handwritten text for forensic document examiners. Proceedings of the 12th Conference of

the International Graphonomics Society; June 26 – 29; Salerno, Italy. International Graphonomics Society, pp. 8–12, http://www.graphonomics.org/publications.php

Found, B., Rogers, D. (2008). The probative character of forensic handwriting examiners' identification and elimination opinions on questioned signatures. Forensic Science International, Vol. 178, pp. 54–60.

Kam, M.; Wetstein, J.; Conn, R. (1994). Proficiency of professional document examiners in writer identification. Journal of Forensic Sciences, Vol. 39, pp. 5–14.

Kam, M.; Fielding, G.; Conn, R. (1997). Writer identification by professional document examiners. Journal of Forensic Sciences, Vol. 42, pp. 778–86.

Kam, M.; Gummadidala, K.; Fielding, G.; Conn, R. (2001). Signature authentification by forensic document examiners. Journal of Forensic Sciences, Vol. 46, pp. 884–8.

Kam, M. (2010). Proficiency testing and procedure validation for forensic document examination. Final Report to the Technical Support Working Group, Requirement number 000/IS-001-DREXEL-03-FP

Sita, J.; Found, B.; Rogers, D. (2002). Forensic handwriting examiners' expertise for signature comparison. Journal of Forensic Sciences, Vol. 47, pp. 1117–24.

Tabachnick, B. G., and Fidell, L. S. (2007). Using Multivariate Statistics , 5th ed. Boston: Allyn and Bacon.

Tabachnick, G. G., and Fidell, L. S. (2007). Experimental Designs Using ANOVA. Belmont, CA: Duxbury.