# Investigating the Potential for Training Context Effects to Influence Forensic Document Examiners' Relative Skill at Writer Individualization and Exclusion

*Tonya Trubshoe,[1] Bryan Found [2] [3]*

**Abstract.** *The relative ability of forensic document examiners (FDEs) to provide support for the proposition of individualization or exclusion on the basis of handwriting features was investigated by surveying opinions expressed in case files by one laboratory's FDEs and comparing this data to blind trial test results taken over a five year period. The survey of FDEs opinions on reports showed that opinions were skewed towards support for writer individualization over writer exclusion 92% of the time. Since historically FDEs develop their skills with respect to individualization/exclusion primarily on case files, it is proposed that this unbalanced training context may skew their abilities to carry out the tasks. To determine one laboratory's capacity to correctly provide both individualization and exclusion evidence, results of blind validation trials were analyzed. For natural writing written and not written by the specimen writer, FDEs were 62 times more inconclusive when providing support for exclusion of the specimen writer when the specimen writer did not author the questioned sample, than they were for providing support for individualization when the specimen writer wrote the questioned sample. An intriguing possibility is that because of the unbalanced training set, government FDEs may acquire skills which are skewed towards individualization over exclusion.*

## 1. Introduction

A major task of FDEs is the examination and comparison of handwriting and signatures for the purpose of expressing an opinion as to whether the forensic findings support the proposition that a known writer did or did not write a questioned sample of handwriting. The expertise associated with the acquisition of the perceptual and cognitive skill set used to carry out forensic handwriting tasks has been shown to be real and demonstrable (Kam, Wetstein and Conn, 1994; Kam, Fielding and Conn, 1997; Found, Sita and Rogers, 1999; Kam, Gummadidala, Fielding and Conn, 2001; Sita, Found and Rogers, 2002). This expertise develops over time, with historical training programs based principally around exposure of a trainee to casework over a three to five year period. This training is mentored by colleagues that have, through a similar process, been deemed qualified.

The historical approach to skill acquisition may appear to be sound, as, "people, and experts in particular, learn from experience - this is one of the important cornerstones of intelligence and expertise" (Dror, 2013). However it is the character of the experience,

1 Document Examination Solutions Pty Ltd., Western Australia, tonya@docexamsolutions.com.au

2 Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, bryan. found@police.vic.gov.au

3 Program in Expertise, Evidence & Law, Faculty of Law, University of New South Wales

and not the experience itself, that is critically important since it is feedback that drives expertise development. Unfortunately, the majority of opinions expressed during the training phase of expertise development are based on casework problems where the ground truth is not known and opportunities for valid feedback are suboptimal. Of interest here is the context within which this skill is learned and develops, and whether this context may be a factor in limiting the relative ability of an FDE to provide valid evidence to support individualization or elimination when comparing a known to a questioned population of writing features.

Relevant to the work described here is the recent explosion of interest in the field now called *Cognitive Forensics*. This field examines a myriad of factors and considerations associated with using human perceptual and cognitive processes as an instrument in the forensic sciences. Although much of the discussion relating to cognitive factors in the literature has been around context information and bias, the field encompasses a much wider array of considerations when dealing with human perceptual and cognitive processes at play in the forensic sciences. This is well summarized by Dror (2015) and includes discussion on issues around bias, human communication, perception of role, cognitive contamination, peer review processes, selection of trainees, and training. Further broad discussions have also been published (Saks, Risinger, Rosenthal and Thompson, 2003; Dror, 2008; Dror, 2011; Thompson, 2011; Busey and Dror, 2011; Dror, 2012; Dror, 2013; Kassin, Dror & Kukucka, 2013; Dror and Stoel, 2014; Stoel, Berger, Kerkhoff and Mattijssen, 2014; Stoel, Dror and Miller, 2014; Found, 2014; Dror, 2015).

The research presented here was conducted in 2005, well before cognitive factors in forensic science became a topic of intense interest. It now fits well within the discussion around cognitive factors as it relates to *Base Rate* information discussed in recent times by both Stoel et al (2014) and Dror (2015) where the environment for case submission can result in an expectation regarding the examination outcome. In this case, a police laboratory is the subject of the research and therefore it might be expected that the evidence submitted would be to some extent incriminating. This study therefore aims to further investigate and characterize any difference between government trained FDEs individualization and exclusionary

opinions within this context. We compared a single government laboratory's FDEs opinion profiles on actual casework, with those profiles on handwriting validation trials that they have undertaken. The experimental rationale for this comparison was that since more training is undertaken within the context of a police laboratory, and since in most instances suspects are identified on evidence other than the handwriting prior to the submission of cases, that the case submission profile will result in a skewed set of writing comparison tasks for FDEs to train on. This may result in a more developed individualization skill than an elimination skill. This potentially may be evidenced in the results of these same FDEs on their blind testing results.

## 2. A review of opinions FDEs expressed in casework

### 2.1 Subjects

Over the five year period investigated, the number of FDEs qualified to give handwriting and signature opinion evidence in the laboratory under study fluctuated between four and five. The experience of these examiners ranged from 3 ½ to 25 years.

### 2.2 Procedures

Case files involving handwriting and signature examinations were recorded from case record books for the five year period to determine:

- The type of examination required (handwriting, signature, or both)
- Whether a statement was produced; and
- The level of and number of opinions expressed in statements

Of the case files received for handwriting and / or signature examination, 342 had opinions expressed in the form of a statement.

The opinion levels used by the FDEs in the cases reviewed are the same opinion levels as were used in 2005 by the Australian Special Advisory Group method (Found and Rogers, 1999) and were defined as:

- The questioned handwriting **was written by** the writer of the handwriting specimens. This is an opinion of *moral certainty* on the part of the examiner. For some FDEs this opinion level translates into *very strong support for the proposition that the questioned handwriting was written by the writer of the handwriting specimens*.

- There are indications that the questioned handwriting **was written by** the writer of the handwriting specimens. For some FDEs this opinion translates to statements such as *moderate support for the proposition that the questioned handwriting was written by the writer of the handwriting specimens*

- **No opinion** can be expressed as to whether or not the questioned handwriting was written by the writer of the handwriting specimens. This is an inconclusive opinion.

- There are indications that the questioned handwriting **was not written by** the writer of the handwriting specimens. For some FDEs this opinion translates to statements such as *moderate support for the proposition that the questioned handwriting was not written by the writer of the handwriting specimens*

- The questioned handwriting **was not written by** the writer of the handwriting specimens. This is an opinion of moral certainty on the part of the examiner. For some FDEs this opinion level translates into *very strong support for the proposition that the questioned handwriting was not written by the writer of the handwriting specimens*.

Alongside these opinions, opinions as to common authorship, where specimen handwriting or signatures could not be provided, were also given. The common authorship opinions recorded also corresponded to the above five opinion levels.

## 2.3 Subjects' performance on blind trials

### 2.3.1 Subjects

Over the five year period investigated, four examiners qualified to give handwriting and signature opinion evidence participated in the handwriting and signature validation trials conducted by the Forensic Expertise Profiling Laboratory, School of Human Biosciences, La Trobe University, Australia. It should be noted that every examiner did not partake in every trial.

### 2.3.2 Materials

The handwriting validation trials were received once a year for examination and consisted of between 100 and 250 handwriting questioned samples to be compared with known handwriting samples of the specimen writer. Details regarding the rationale and structure of these trials has been published (Found and Rogers, 1999; Found and Rogers, 2003).

The subjects were blind to the writers of the questioned and specimen samples. The questioned handwriting samples consisted of a random selection of the following writings

- Natural writing written by the specimen writer
- Natural writing not written by the specimen writer
- Disguised writing written by the specimen writer
- Disguised writing not written by the specimen writer
- Forged writing not written by the specimen writer

For each trial, the questioned and specimen handwriting samples were of the same text and format as one another and utilized the same type of ball-point pen and make of paper. Samples were scanned at 600 pixels per inch and printed on an inkjet printer. Each year's trial was comprised of a different section of text repeated.

| Year | Was written by | Indications was written by | Inconclusive | Indications was not written by | Was not written by |
|------|---------------|---------------------------|--------------|-------------------------------|--------------------|
| 2000 | 17 | 21 | 36 | 0 | 2 |
| 2001 | 20 | 21 | 24 | 0 | 0 |
| 2002 | 29 | 17 | 12 | 0 | 1 |
| 2003 | 32 | 15 | 6 | 2 | 3 |
| 2004 | 26 | 21 | 14 | 3 | 1 |
| Total number of opinions | 124 | 95 | 92 | 5 | 7 |
| % of total opinions expressed by type | 38.4 | 29.4 | 28.5 | 1.6 | 2.2 |

Table 1: The number of handwriting opinions for each opinion type

| Year | Individualization | Exclusion |
|------|-------------------|-----------|
| 2000 | 38 | 2 |
| 2001 | 41 | 0 |
| 2002 | 46 | 1 |
| 2003 | 47 | 5 |
| 2004 | 47 | 4 |
| Total number of opinions | 219 | 12 |
| % of opinions by type | 94.8 | 5.2 |

Table 2: The number of handwriting opinions in support of individualization compared to exclusion.

### 2.3.3 Procedures

When receiving the sample and answer booklets, the examiners were informed that:

- The questioned samples were a combination of the different writing types explained in section 2.3.2.
- The questioned and specimen samples were written over the same time period.
- The questioned and specimen samples were to be compared and an opinion expressed in the answer booklet.
- The same opinion levels as described above were used.

## 3. Results

### 3.1 Subjects' opinion profiles on casework

In total, 530 opinions were formed in the examination of the 342 cases involving handwriting and/or signatures over the five year period. More opinions were given than cases examined as some cases involved multiple writings as well as types and styles of writing and signatures. Of the 530 opinions, 323 were opinions on the authorship of handwriting and 207 were opinions on the authorship of signatures.

### 3.2 Results for handwriting opinions

Table 1 shows the number of handwriting opinions expressed, for each of the five opinion types over the five year period. Of the opinions given, on case files received and examined, an unqualified opinion in support of individualization occurred 38 percent of the time and a qualified opinion in support of individualization occurred 29 percent of the time. A qualified or unqualified opinion in support of writer exclusion however, was only given two percent of the time for each opinion type. A total of 29 percent of all examinations undertaken resulted in no opinion being formed.

| Year | Was written by | Indications was written by | Inconclusive | Indications was not written by | Was not written by |
|---|---|---|---|---|---|
| 2000 | 8 | 8 | 11 | 4 | 0 |
| 2001 | 6 | 13 | 18 | 2 | 0 |
| 2002 | 18 | 13 | 22 | 3 | 0 |
| 2003 | 20 | 11 | 5 | 2 | 5 |
| 2004 | 21 | 4 | 11 | 1 | 1 |
| Total number of Opinions | 73 | 49 | 67 | 12 | 6 |
| % of total opinions expressed by type | 35.3 | 23.7 | 32.4 | 5.8 | 2.9 |

Table 3: The number of signature opinions for each opinion type

| Year | Individualization | Exclusion |
|---|---|---|
| 2000 | 16 | 4 |
| 2001 | 19 | 2 |
| 2002 | 31 | 3 |
| 2003 | 31 | 7 |
| 2004 | 25 | 2 |
| Total number of opinions | 122 | 18 |
| % of opinions by type | 87.1 | 12.9 |

Table 4: The number of signature opinions in support of individualization compared to exclusion

Of the 323 handwriting opinions that resulted in statements being issued, 231 were in support of individualization or exclusion opinions and 92 were inconclusive. Table 2 shows that, of the 231 opinions given, 95 percent were in support of individualization while only five percent were for exclusion.

### 3.3 Results for signature opinions

Table 3 shows the number of signature opinions given for each of the five opinion types, over the five year period. Of the opinions given, on case files received and examined, an unqualified opinion in support of individualization occurred 35 percent of the time and a qualified opinion in support of individualization occurred 24 percent of the time. A qualified or unqualified opinion in support of exclusion however, was only given six and three percent of the time for each opinion type respectively. A total of 32 percent of all examinations undertaken resulted in no opinion being offered.

Of the 207 signature opinions that resulted in statements being issued, 140 were for individualization or exclusion opinions and 67 were inconclusive. Table 4 shows that, of the 140 opinions given, 87 percent were for individualization while only 13 percent were for exclusion.

### 3.4 Summary of results for handwriting and signature opinions

Of the 371 handwriting and signature examinations that resulted in opinions being given, 341 (92%) were in support of individualization and 30 (8%) for exclusion (Table 5).

| Year | Individualization | Exclusion |
|---|---|---|
| 2000 | 54 | 6 |
| 2001 | 60 | 2 |
| 2002 | 77 | 4 |
| 2003 | 78 | 12 |
| 2004 | 72 | 6 |
| Total number of opinions | 341 | 30 |
| % of opinions by type | 91.91 | 8.09 |

Table 5: The number of handwriting and signature opinions in support of individualization compared to exclusion

| | Opinions on specimen writer samples | | | Opinions on samples not by the specimen writer | | |
|---|---|---|---|---|---|---|
| | Unqualified | Qualified | Inconclusive | Unqualified | Qualified | Inconclusive |
| FDE 1 | 56 | 1 | 0 | 0 | 12 | 25 |
| FDE 2 | 79 | 5 | 2 | 0 | 46 | 110 |
| FDE 3 | 61 | 5 | 2 | 0 | 0 | 97 |
| FDE 4 | 84 | 7 | 0 | 0 | 34 | 153 |
| Total | 280 | 18 | 4 | 0 | 92 | 385 |

Table 6: Handwriting validation trial results for All FDEs

The results of the case file review, given the markedly skewed distribution of opinion profiles for individualization compared with exclusion opinions, provides strong support for the proposition that individuals are exposed to significantly more tasks resulting in individualization opinions than tasks resulting in exclusion opinions. Although there is no way, based on current empirical evidence, to assess whether this skewed training environment does result in a skill weighted toward individualization, the data does show that at least the potential for this effect to be true is real.

## 4. Subjects' opinion profiles on blind trials

Answers on the blind validation trials were marked by the Forensic Expertise Profiling Laboratory as correct, misleading or inconclusive. In this set of trials, no misleading opinions were expressed. Of the four different writing types making up the questioned samples on the trials, only the answers given in relation to natural writing, written by and not written by the specimen writer, were relevant, and used, in this research investigation.

## 5. Summary of results for all FDEs' opinions on the validation trials

Over the five year period, 779 samples relating to the individualization and exclusion of writers in handwriting validation trials were examined by the four examiners qualified to do so. Of these samples, 302 involved natural writing written by the specimen writer (individualization examinations) and 477 involved natural writing not written by the specimen writer (exclusion examinations). Table 6 shows the number of opinions given for the four examiners over the five year period. Of the possible 302 individualizations possible, 298 FDEs' opinions were correct (280 unqualified and 18 qualified opinions). Of the possible 477 eliminations possible, 92 qualified opinions were given. Therefore, FDEs did not provide
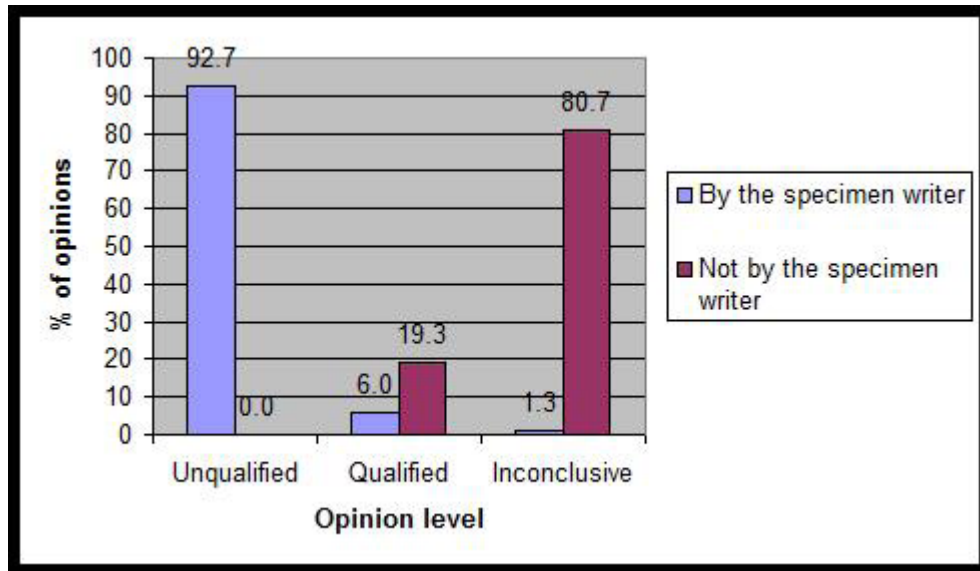
Figure 1: Opinion level and percentage of opinions given for the individualization and exclusion samples on all trials by all FDEs combined.

opinions on four of the individualization samples of the 302 provided and 385 of the exclusion samples of the 477 provided.

Figure 1 is a schematic representation of the data provided in Table 6 in the form of percentage of opinions for each opinion type for natural writing written by the specimen writer (individualization examinations) and natural writing not written by the specimen writer (exclusion examinations) for the handwriting validation trials examined.

As can be observed, no unqualified exclusion opinions were expressed by the FDEs on the samples of handwriting known not to have been written by the specimen writer, whereas 92.7 percentage of the samples known to have been written by the specimen writer fell into this unqualified correct category. In addition, only 1.3 percentage of samples written by the specimen writer attracted an inconclusive result as compared to 80.7 percentage of inconclusive opinions on samples not written by the specimen writer. This means that these FDEs were 62 times more inconclusive on exclusion samples as compared to individualization samples on these blind trials.

## 6. Discussion

Training based on casework is an important element during the document examiner's three to five year internal training process. However, as we have shown, the training set is significantly skewed. It was found from the examination of case files that of the opinions given, 95 percent were for individualization opinions while only five percent were associated with exclusion. Analysis of blind trial results on this same group of FDEs confirmed that their ability to correctly provide evidence in support of the exclusion of writers was significantly less than their ability to individualize writers. From the blind trial results these FDEs were 62 times more inconclusive on exclusion samples as compared to individualization samples on the blind trials.

Training context is an intriguing candidate for FDEs' differential ability at individualization and exclusion tasks. This is, however, not the only potential source of the effect. Logically, and for most practical case file tasks, if the questioned writing sample is similar to the suspects writing sample then there are only three propositions to explain the observation; the suspect wrote the sample, the questioned sample is a forgery of the suspects writing, or a chance match has occurred between the suspects writing and the questioned writing. The chance match possibility, given an adequate sample of writing, is low due to the extent of inter-writer variation in the population. The difficulty that individuals have at forging the handwriting characteristics of others is well

understood and the chance of forging handwriting without leaving evidence of the process behind is low (for a sufficiently complex population of handwriting features). This can leave the FDE to support the proposition that the suspect writer wrote the questioned sample. When dissimilarities are detected between the questioned writing and the suspect sample, there are a greater number of propositions for the FDE to address. The questioned sample may be different from the specimen sample for any one of the following reasons; the suspect may not have written the questioned sample, the suspect's writing may not have been sampled fully, the suspect may have disguised his/her writing, or the suspect may have more than one style of normal handwriting. Since there is a greater number of propositions that could be advanced to explain why a questioned sample is dissimilar than there is to explain why a questioned sample is similar, this may result in a greater proportion of inconclusive opinions regarding potential exclusion examinations.

With regards to this research, the differing number of propositions to explain similarity or dissimilarity cannot be directly investigated as to its impact on FDEs differential individualization and exclusion skills without first controlling for the training context issue described.

## 7. Methods to control for any potential context effect in the training and casework environment

Clearly, the most appropriate approach to the base rate issue described in this research is to reconfigure training approaches such that trainees would have close to equal opportunities to provide individualization and exclusionary opinions. Coupled with this would be a training environment where the ground truth of the problem being worked is known. Casework does not provide this opportunity. However, the development of ground truth known, case databases could enable examiners to apply their expertise in an environment where feedback would be available. In addition, Stoel et al. (2014) suggest adding known outcome false cases to the workflow (which in this case would assist in balancing the base rate). Although these authors concede that "the direct numerical effect of these false cases on the base rate may be relatively small … the psychological effect of the presence of false cases may

… be larger." This, of course, does not detract from continued mentored working on actual cases.

Through the methods described above, the skewed training set described here can be compensated for by the regular examination of balanced individualization and exclusion practical problems. The medium to long term effect of this training approach can then be monitored and assessed through the performance of new practitioners on future blind trials. Although it is the case that the skewed case file profile is unlikely to change, the confidence that FDEs have in correctly providing opinions regarding individualization and exclusion of writers may improve. This in turn should provide at least the potential for an improved understanding of the character of the evidence within the forensic science framework and a more balanced forensic service to both the courts and to those individuals suspected of committing a crime.

## References

Busey, T., Dror, IE. (2011). The Fingerprint Sourcebook. McRoberts, A, (ed.) National Institute of Justice: Washington, D.C., US. Chapter 15. Special Abilities and Vulnerabilities in Forensic Expertise, 3 -23.

Dror, IE. (2008). Biased Brains. Police Rev, 116, 20-23.

Dror, IE. (2011). Paradoxical Brain. Kapur N (ed.). Cambridge University Press. The paradox of human expertise: why experts get it wrong, 177-188.

Dror, IE. (2012). Combating bias: The next step in fighting cognitive and psychological contamination. J Forensic Sci. 57, 276-277.

Dror, IE. (2013). Practical solutions to cognitive and human factor challenges in forensic science. Forensic Sci Policy Manag. 4, 1-9.

Dror, IE., Stoel, RD. (2014). Cognitive Forensics: Human cognition, contextual information, and bias. Encyclopaedia of Criminology and Criminal Justice. Springer Publishing.

Dror, IE. (2015). Cognitive neuroscience in forensic science: understanding and utilizing the human element. Phil. Trans. R. Soc. B. 370;20140255:http//dx.doi.org/10.1098/rstb.2014.0255.Found, B., Rogers, D. (1999). Documentation of forensic handwriting comparison and identification method: A modular approach. J Forensic Doc Exam 12, 1-68.

Found, B., Sita, J. & Rogers, D. (1999). The development of a program for characterizing forensic handwriting examiners' expertise: Signature examination pilot study. Journal of Forensic Document Examination, Volume 12.

Found, B., Rogers, D. (2003). The initial profiling trial of a program to characterise forensic handwriting

examiners' skill. Journal of the American Society of Questioned Document Examiners, 6, 72-81.

Found, B. (2014). Deciphering the human condition: the rise of cognitive forensics. Aust J Forensic Sci.

Kam, M., Wetstein, J., & Conn, R. (1994). Proficiency of professional document examiners in writer identification. J Forensic Sci, 39, 5-14.

Kam, M., Fielding, G., & Conn, R. (1997). Writer identification by professional document examiners. J Forensic Sci 42, 778-786.

Kam, M., Gummadidala, K., Fielding, G., & Conn, R. (2001). Signature authentication by forensic document examiners. J Forensic Sci 46, 884-8.

Kassin, SM., Dror, IE., & Kukucka, J. (2013). The forensic conformation bias: Problems, perspectives, and proposed solutions. J Appl Res Mem Cogn. 2, 42-52.

Saks, MJ., Risinger, DM., Rosenthal, R., & Thompson, WC. (2003). Context effects in forensic science: A review and application of the science of science to crime laboratory practice in the United States. Science and Justice, 43, 2, 77-90.

Sita, J., Found, B., & Rogers, D. (2002). Forensic handwriting examiners' expertise for signature comparison, J Forensic Sci 47,5, 1117-1124.

Stoel, RD., Berger, CEM., Kerkhoff, W., Mattijssen, EJAT., & Dror, IE. (2014). Minimising contextual bias in forensic casework. In M. Hickman and K. Strom (Eds.), Forensic Science and the Administration of Justice. SAGE Publishing.

Stoel, RD., Dror, IE., & Miller, LS. (2014). Bias among forensic document examiners: Still a need for procedural changes. Aust J Forensic Sci. 46, 91-97.

Thompson, W. (2011). What role should investigative facts play in the evaluation of scientific evidence? Aust J Forensic Sci. 43, 123-134.