NISTIR 8282R1

Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach

The Expert Working Group for Human Factors in Handwriting Examination

Updated May 2021

NIST National Institute of Standards and Technology

NIJ | National Institute of Justice
STRENGTHEN SCIENCE. ADVANCE JUSTICE.

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8282r1

NISTIR 8282R1

Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach

The Expert Working Group for Human Factors in Handwriting Examination

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8282r1

Updated May 2021

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Interagency or Internal Report 8282R1
Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8282R1, 253 pages (May 2021)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8282r1

In Memoriam
This report is dedicated to the memory of Dr. Bryan Found, a valued contributor to this project and a friend who is dearly missed.

This page intentionally left blank.

Table of Contents

List of Tables

List of Figures

This page intentionally left blank.

Glossary

A

Accuracy: Similar to validity in that it relates to correctness of a result (i.e., closeness of measurements/outcomes to the true value).

Alignment: Position of writing with respect to a real or imaginary baseline.[1]

Allograph: Different forms of the same letter (or grapheme), such as capital hand-printed "A" and cursive "a."[2]

Arrangement: An element of handwriting style relating to the placement of text on the page that includes characteristics such as margin habits, interline and inter-word spacing, indentations, and paragraphing.[3]

Authentic: When a document/handwriting is genuine.[4]

Authorship: Origin of the content of a document. *See also* Writership.

B

Baseline: The real or assumed line upon which handwriting is produced.[5]

Bias: A systematic pattern of deviation.

Blind Case: A case that has been developed with the intention of testing the examiner or the examination process and in which the ground truth is known. Critically, the examiner is not aware the case is not genuine.

Blind Declared Case: Blind cases the examiner knows will be inserted into routine casework. The examiner will not know which cases are blind.

Blinding: Systematically shielding an examiner from task-irrelevant contextual information.

C

Chance Match: The occurrence of naturally produced handwriting by two different writers that displays the same handwriting characteristics such that the writing cannot be distinguished.[6]

Character: Letters, numbers and symbols; graphemes.[7]

---

[1] R. A. Huber and A. M. Headrick, *Handwriting Identification: Facts and Fundamentals* (Boca Raton: CRC Press LLC, 1999), 394.

[2] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*.

[3] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 91.

[4] Bryan J. Found and Carolyne Bird, "The Modular Forensic Handwriting Method—2016 Version," *Journal of Forensic Document Examination* 26 (2016): 71, https://doi.org/10.31974/jfde26-7-83.

[5] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[6] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[7] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

Character Set: A standard set of letters (basic written symbols or graphemes), which is used to write one or more languages based on the general principle that the letters represent phonemes (basic significant sounds) of the spoken language or other symbols that convey meaning.[8]

Characteristic: A feature, quality, attribute, or property of writing.

Class: The handwriting characteristics shared by a group of writers; for example, copybook writing.[9]

Cognitive Bias: A systematic pattern of deviation in human judgement.

Collected Writing: A subset of known writing. Samples of a known person's handwriting/signatures that have been produced throughout the course of day-to-day business, are typically not related to the case at hand, and have been collected by the case submitter for the purposes of comparison against questioned material. Examples include letters, diaries, business records, forms, or checks. These can also be known as normal course specimen or course-of-business specimens.[10]

Commencement and Termination Strokes: Strokes at the beginning or end of characters that lead into or out of the letter.

Common Writership: A comparison of handwriting where the forensic document examiner (FDE) is asked to give an opinion on whether a group of questioned documents have been produced by the same writer.[11] *See also* Intra-comparison.

Comparable: The attribute of being suitable for comparison; for example, handwriting in the same style.[12]

Complexity: A combination of speed, skill, style, and construction that contributes to handwriting being difficult to simulate.[13]

Connecting Stroke: A line adjoining two adjacent characters.[14]

Connections: The union of two characters; for example, in cursive writing.[15]

Consistent: Similar, regular throughout a passage of writing or between multiple signatures.[16]

[8] Adapted from Wikipedia's entry for "alphabet."

[9] J. S. Kelly and B. S. Lindblom, eds., *Scientific Examination of Questioned Documents*, 2nd ed. (Boca Raton: CRC Press—Taylor & Francis Group, 2006), 409.

[10] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 71.

[11] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[12] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[13] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[14] *Standard Terminology Relating to the Examination of Questioned Documents,* ASTM E2195-02e1, (West Conshohocken, PA: ASTM International, 2003).

[15] ASTM International. *Standard Terminology Relating to the Examination of Questioned Documents*.

[16] ASTM International. *Standard Terminology Relating to the Examination of Questioned Documents*.

Construction: How a character, word, or signature has been produced, including number, direction, and sequence of strokes.[17]

Contemporaneous Writing: Two or more samples of writing that were written within a similar time period.

Context: The set of circumstances or facts that surround a case.

Context-Manager Model: A type of contextual information management procedure whereby a forensic expert or administrator filters discipline- and task-irrelevant contextual information from the examiner who is to perform the examination.

Contextual Bias: A type of cognitive bias to denote human judgement being influenced by irrelevant contextual information.

Contextual Information: Knowledge, whether relevant or irrelevant, concerning a particular fact or circumstance related to a case or examination. Contextual information is conceptualized in different levels (see sections 2.1.2 to 2.1.6). These levels are ordered with respect to how far removed the information is from the questioned material and the examination.

Contextual Information Management (CIM): Actions to optimize the flow of information to and from a forensic expert to minimize the potential for contextual bias.

Copybook Systems: A particular manual of writing instruction that provides model letter designs for the student to copy.[18]

D

Diacritic: A mark used with a letter or group of letters to indicate a sound value that is different from that of the letter(s) without it. Often incorrectly used to describe the "i" dot.[19]

Difference: Consistent, repeated dissimilarity in a structural or line quality feature, generally not observed as natural variation in one writer.[20] May be referred to as a significant or fundamental difference.

Dimensions: The physical measurements or size of writing, particularly the absolute size, horizontal and vertical measures, and proportions.[21]

Disguised Writing: Deliberately altered writing.[22]

Dissimilarity: A pictorial, line quality, or structural feature present in a body of writing but not observed in the same form in a compared body of writing.[23]

---

[17] ASTM International. *Standard Terminology Relating to the Examination of Questioned Documents*.

[18] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 398.

[19] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 114.

[20] *Standard Guide for Examination of Handwritten Items,* ASTM E2290-03, (West Conshohocken, PA: ASTM International, 2003).

[21] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 101–02.

[22] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 71.

[23] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 27.

Distorted Writing: Writing that does not appear to be natural but might be natural. This appearance can either be caused by voluntary factors (e.g., disguise or simulation) or involuntary factors (e.g., physical condition of the writer or writing conditions).[24]

Document: Any material containing marks, symbols, or signs visible, partially visible, or invisible (to the naked eye) that may ultimately convey meaning or a message.[25]

E

Embellishments: Flourishes, ornaments, or underscores.[26]

External (Extrinsic) Factors: Writing conditions like underlying writing surface, substrate, writing implement, writing position, or interruptions during the writing activity that affect the handwriting movement or the resulting writing.

F

Feature: An aspect of a character or the handwriting in general.[27]

Flourish: An ornamental or exaggerated pen stroke.[28]

Fluency: The speed and skill level of the writing.[29]

Forensic Discipline: A specialized branch or field of forensic science (e.g., handwriting examination, DNA analysis, latent print examination, and bloodstain pattern analysis).

Forensic Document Examiner (FDE): An examiner trained in the various examination types comprising the field of forensic document examination, including analyses or comparisons of handwriting, print process, ink, indented impressions, and paper. Note that in some countries the term forensic handwriting examiner refers to an examiner of handwriting, and the term FDE is used for examiners of all other areas encompassed by the broad term forensic document examination.

G

Grapheme: The abstract concept of a letter of the alphabet.[30]

Guidelines: Lines that show a route to follow when simulating handwriting or signatures. These can exist in the form of pencil lines or indentations or can be created by the use of transmitted light shone through a document containing the entries to be copied.[31]

---

[24] ASTM International. *Standard Guide for Examination of Handwritten Items*.

[25] Kelly and Lindblom, *Scientific Examination of Questioned Documents*, 411.

[26] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 115.

[27] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 71.

[28] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[29] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[30] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 401.

[31] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

H

Handwriting or Writing: Writing in any form (such as cursive writing, hand printing, signatures, numbers). Although "handwritten" is used as a general term, writing may not be produced using the hand but may be the result of some other part of the body (e.g., mouth or foot) directly manipulating a writing or marking instrument.[32]

I

Inconclusive Opinion: An opinion expressed when a handwriting examination has been undertaken, but the FDE is unable to make a determination with regard to writership; for example, because of the presence of both similarities and dissimilarities.

Indented Impressions: Markings or imprints on the paper surface caused by the pressure of a writing instrument on the pages or paper above.[33]

Insufficient Opinion: A determination made by an FDE that the material to be examined does not contain enough information for an examination to be conducted. This may be because of the amount, complexity, comparability, line, reproduction, or writing quality of the material. In many instances, FDEs report an inconclusive opinion, explaining limitations/insufficiency, rather than reporting an insufficient opinion.

Inter-comparison: Comparison of two or more bodies of writing to determine whether they have been written by more than one writer.

Internal (Intrinsic) Factors: Conditions such as age, illness, disease, fatigue, emotional state, medication, or intoxication by drugs or alcohol that affect the handwriting movement and the resulting writing.

Intra-comparison: Comparison of handwriting within one document or purportedly by one writer, to determine whether the handwriting has been written by one person.[34]

Irrelevant Information: Information that is not pertinent or applicable to the subject, material, or question being considered. The consideration may be broad (i.e., discipline level) or specific (i.e., task level).

K

Known Writing (also K, Exemplar or Standard): Writing of established origin associated with the matter under investigation.[35] Known writing may be collected course-of-business documents or—if written for the purpose of comparison—requested, witnessed, or dictated.

[32] ASTM International. *Standard Guide for Examination of Handwritten Items*.

[33] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 71.

[34] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 72.

[35] ASTM International. *Standard Guide for Examination of Handwritten Items*.

L

Laboratory: For the purposes of this report, an agency, team, or sole practitioner who provides a forensic document examination service.

Legibility or Writing Quality: Ease of recognition of letters.[36]

Limitation: A constraint to the examination, comparison, or opinion formation process (e.g., non-original documents, limited quantity of material).[37]

Line Continuity: Continuity of the writing line. Discontinuity may be in the form of pen lifts, pen stops or hesitations, or retouching of characters to improve pictorial appearance or legibility.[38]

Line Quality: The degree of regularity of handwriting, resulting from a number of factors, including speed, skill, freedom of movement, execution rhythm, and pen pressure. May vary from smooth and fluent to tremulous and erratic.[39]

Linear Sequential Unmasking (LSU): A type of CIM procedure that specifies the optimal order in which forensic experts should examine the unknown material (e.g., questioned writing) and reference material (e.g., known writing) to conduct a comparison. The experts must examine and document the unknown material before being exposed to the reference material, therefore working from the evidence to the suspect.[40] The term LSU has been coined by Dror and colleagues[41] to stress that the examiner is not allowed unlimited back and forth access between the questioned and known material. LSU follows the same basic principles of sequential unmasking; however, it also requires FDEs to specify a level of confidence in their opinion regarding the material under examination.[42]

---

[36] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*.

[37] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 72.

[38] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 118.

[39] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 120.

[40] D. E. Krane et al., "Sequential Unmasking: a Means of Minimizing Observer Effects in Forensic DNA Interpretation," *Journal of Forensic Sciences* 53, no. 4 (Jul 2008), https://doi.org/10.1111/j.1556-4029.2008.00787.x.

[41] I. E. Dror et al., "Letter to the Editor—Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making," *Journal of Forensic Sciences* 60, no. 4 (Jul 2015), https://doi.org/10.1111/1556-4029.12805, https://www.ncbi.nlm.nih.gov/pubmed/26088016. "Sequential unmasking allows unlimited and unrestricted changes to the evidence once exposed to the reference material. We believe it is important to impose limits and restrictions for when examiners are permitted to revisit and alter their initial analysis of trace evidence. The analysis of traces is most objective when the examination is 'context free'—that is, prior to exposure to the known reference samples. However, seeing the reference samples could alert the examiner to a possible oversight, error, or misjudgment in the analysis of the trace evidence. Here, we seek to strike a balance between restrictive procedures that forbid analysts from changing their opinion and those that allow unlimited and unrestricted changes. The requirement that changes be documented does not eliminate the possibility that such changes arose from bias—it only makes that possibility more transparent."

[42] Because the features that must be considered in a handwriting case are generally not defined before the case, taking a strict approach to LSU in handwriting examination could result in a loss of evidential strength. This is discussed more in section 2.1.3.

N

Natural Variation: Those deviations among repetitions of the same handwriting characteristic(s) that are normally demonstrated in the habits of each writer.[43]

No Conclusion: An opinion expressed when no opinion regarding authorship can be drawn because of insufficient material or the presence of both similarities and dissimilarities (i.e., either an Inconclusive or Insufficient Opinion).

Non-original: Reproduction of a document; for example, photocopied, faxed, scanned, or photographed.[44]

Normal Writing (also Natural Writing): Any specimen of writing executed without an attempt to control or alter its usual quality of execution.[45]

P

Pen Direction: The direction the pen moves to produce a character, connection, or signature.[46]

Pen Lift: An interruption in a stroke caused by removing the writing instrument from the writing surface.[47]

Proportions: Relative size of characters and elements of characters (e.g., of bowl to staff in "d"). May also refer to the relative size of words.[48]

Proposition: A statement or outcome to be tested during examination. There are generally two opposing propositions to be tested: (1) The same writer produced A and B or (2) different writers produced A and B.[49]

Q

Quality: See Legibility or Writing Quality, Line Quality, and Reproduction Quality.

Questioned Writing: Handwriting about which the authenticity or writership is in doubt. Sometimes referred to as Q writing.[50]

[43] *SWGDOC Standard for Examination of Handwritten Items,* Version 2013-1, (Scientific Working Group for Forensic Document Examination (SWGDOC), 2013).

[44] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 72.

[45] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[46] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[47] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[48] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 102.

[49] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 72.

[50] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

R

Random Error: A component of error whereby replicate measurements vary in an unpredictable way. Sources of random error are usually unexplained and are therefore difficult to control.[51]

Range of Variation: The extent to which the writing habits of an individual are reproduced, or vary, on repeated occasions. Variation may occur in any of the handwriting characteristics, from the construction of letters and numbers to slant, alignment, and line quality.

Relevant Information: Information that is pertinent and applicable to the subject, material, or question being considered. The consideration may be broad (i.e., case or discipline level) or specific (i.e., task level).

Reliability: To what degree do single or multiple FDEs reach the same answer under specified tasks and constant conditions. Reliability is related to the degree of random error of the instrument/method, which can include the FDE. The smaller the amount of random error, the more reliable the instrument/method, and vice versa. Two ways to assess reliability are repeatability and reproducibility.[52]

Repeatability: A measure of reliability using the same FDE and the same instrument/method under exactly the same conditions to arrive at the same conclusion or result.

Reproducibility: A measure of reliability using different FDEs and/or differing conditions with the same measurement instrument/method to arrive at the same conclusion or result.

Reproduction Quality: The degree to which a non-original document accurately replicates the features of the original document.

Requested Writing: Handwriting samples written by a particular person specifically for the purpose of comparison to questioned material (as requested by a submitting party).[53]

Retouching: To add lines or strokes to correct, improve, or alter writing.[54]

S

Signature Style: Can be (1) text-based (all allographs legible), (2) mixed style (two or more allographs are legible), or (3) stylized (one or no allographs are legible).[55]

Similarities: Having mutual resemblance and a number of features in common.[56]

---

[51] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[52] For application of the concepts discussed under reliability to forensic science, see B. T. Ulery et al., "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners," *PLoS One* 7, no. 3 (2012), https://doi.org/10.1371/journal.pone.0032800.

[53] Ulery et al., "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners."

[54] Ulery et al., "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners."

[55] L. Mohammed, B. Found, and D. Rogers, "Frequency of Signature Styles in San Diego County," *Journal of the American Society of Questioned Document Examiners* 11, no. 1 (2008).

[56] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 72.

Simplistic Writing: Characterized by non-complex characters or strokes.[57]

Simulation: An attempt to copy or reproduce handwriting.[58]

Skill: How well an individual is able to produce and repeat the formation of handwritten characters.[59]

Slant or Slope: The angle or inclination of the axis of letters relative to the baseline.[60]

Spacing: The distance between characters, words, or lines in writing.[61]

Speed: How fast the writing is produced.[62]

Structural Features: Features relating to the construction of handwriting (e.g., number, position, order, and direction of strokes).[63]

Style (also Design): The general category of allograph (letter form) that is employed to execute writing; for example cursive or hand printing.[64]

Substrate: The material that is written on, usually paper.[65]

Suitability: Sufficient quantity, quality, and complexity specifically for comparison.

Systematic error: A component of error whereby replicate measurements remain constant or vary in a predictable way—for example an uncalibrated instrument would produce a constant systematic error.[66]

T

Task: A piece of work to be undertaken.

Termination Stroke: The final stroke of a character or word.[67]

Tracing: Writing that is created by placing a model underneath the paper to be written on, such that the model can be observed through the paper to provide guidelines to assist in copying.[68]

---

[57] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[58] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[59] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[60] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 408.

[61] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 73.

[62] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[63] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[64] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 95.

[65] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 73.

[66] "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)," updated April 29, 2017, https://jcgm.bipm.org/vim/en/

[67] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

[68] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

Tremor: A lack of smoothness in the writing trace, due to lack of skill, deliberate control of the writing implement, or involuntary movement (e.g., illness).[69]

Turning Points: Position at which a pen line changes direction.[70]

U

Unnatural Writing: A writing movement not typical to day-to-day writing that may be the result of intent, internal, or external factors. Unnatural writing is seen when a person is trying to disguise his or her own writing or trying to simulate that of another writer. Some characteristics of unnatural writing movements include slow speed, poor line quality, poor line continuity with stops or hesitations in the pen line, and blunt commencement and termination strokes.[71]

V

Validity: To what degree do single or multiple FDEs reach the correct answer under specified tasks and constant conditions. A test is valid if it measures what it is supposed to measure.[72] A measure can be reliable and not valid but not vice versa. In other words, reliability is necessary but not sufficient for validity, and if a measurement instrument/method is valid, it is also reliable.

Variation: Having one or more forms (constructions) of a character or word in a naturally produced sample of handwriting.[73]

W

Writer: The physical executor of the handwriting, the person who put "pen to paper."

Writership: Origin of the physical handwriting on a document.[74] *See also* Authorship.

Writing Implement: Any tool used to create a handwritten marking on a substrate. Typically used to describe the use of a pen, pencil, marker, or crayon to create words on paper.[75]

---

[69] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

[70] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

[71] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

[72] See D. Borsboom, G. J. Mellenbergh, and J. van Heerden, "The Concept of Validity," *Psychological Review* 111, no. 4 (Oct 2004), https://doi.org/10.1037/0033-295X.111.4.1061.

[73] Borsboom, Mellenbergh, and van Heerden, "The Concept of Validity."

[74] The term "author" often refers to the creator of the content of writing. Thus, studies have examined who composed the specific essays in *The Federalist Papers* (Alexander Hamilton, James Madison, and John Jay, *The Federalist: A Collection of Essays Written in Favour of the New Constitution as Agreed Upon by the Federal Convention* (1788), https://www.loc.gov/item/09021562/.) that appeared under the pseudonym of "Publius" and who wrote the works attributed to Shakespeare. "Authorship" in that sense is the subject of forensic linguistics (see, for example, R. Zheng et al., "Authorship Analysis in Cybercrime Investigation," in *Intelligence and Security Informatics*, ed. H. Chen, R. Miranda, D. D. Zeng, C. Demchak, J. Schroeder and T. Madhusudan (Berlin, Heidelberg: Springer, 2003).). Because the writer of a physical text might not have been the original author, the Working Group uses the more precise term "writership" throughout this report, rather than the broader term "authorship," to denote the physical executor of the handwriting under examination.

[75] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 73.

Writing Movement: A characteristic of writing seen in letter constructions and connecting strokes that relates to the predominant action of the writing instrument. These movements may be (1) garlanded, where counterclockwise movements predominate; (2) arched, with predominately clockwise movements; (3) angular, where straight lines take precedence to curves; or (4) indeterminable, where the predominating movement is uncertain.[76]

Writing Surface: The underlying surface that a substrate (e.g., paper) is placed on while handwriting is produced. This will impact the pictorial qualities of the writing and can impose a limitation on comparisons.[77]

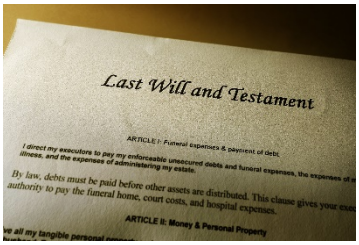
[76] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 131.

[77] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*.

This page intentionally left blank.

Introduction

For 6,000 years, people have made an indelible mark on history with handwriting—the loops, strokes, and other characters that constitute the written form of language. Whether it is a stylus moving across wet clay or a pen moving across paper, handwriting has always been a familiar and idiosyncratic form of expression. The study of handwriting is also an important part of forensic science. By analyzing the characteristics of a handwritten note or signature—not only the slant of the writing and how letters are formed but more subtle features too—a trained forensic document examiner (FDE)[78] may be able to extract valuable information to determine whether a note or signature is genuine and the likely identify of the writer.



Images Courtesy of Fotolia

The results of forensic document examination can have far-reaching consequences that affect a person's life and liberty. An FDE may be called on in a court of law to answer—or to supply information that would help a judge or jury answer—questions involving authenticity and writership. However, several recent studies cited throughout this document highlight the increased recognition and concern that the nature of evidence and human factors have the potential to inadvertently influence forensic examinations, including handwriting examination.

The study of human factors examines interactions between people and the other elements of a system—technology, training, decisions, products, procedures, workspaces, and the overall environment—with the goal of improving both human and system performance. Inadequate training, extraneous knowledge about the suspects in the case or other matters, poor judgement, vision limitations, complex technology, and stress are a few of the factors that contribute to errors. Furthermore, poor management, insufficient resources, and substandard working conditions can also prove detrimental to an examination. Analyzing human factor issues in handwriting examination—how they arise and how they can be prevented or mitigated—can inform the development of strategies to reduce the likelihood and impact of errors.

The National Institute of Justice (NIJ) Office of Investigative and Forensic Sciences (OIFS) and the National Institute of Standards and Technology (NIST) Special Programs Office sponsored the work of the Expert Working Group for Human Factors in Handwriting Examination to encourage and enhance efforts to apply human factors research, reduce the risk of error, and improve the practice of forensic document examination.

[78] For the purposes of this report, both forensic handwriting examiners and forensic document examiners will be referred to as an FDE.

The Expert Working Group for Human Factors in Handwriting Examination

The Expert Working Group for Human Factors in Handwriting Examination (the Working Group) convened in June 2015, the second in a series of expert groups examining human factors in forensic science. It followed a successful and widely read report on human factors in latent print examination (LPE).[79]

The Working Group was charged with conducting a scientific assessment of the effects of human factors on forensic handwriting examination with the goal of recommending strategies and approaches to improve its practice and reduce the likelihood of errors. A scientific assessment, as defined by the Office of Management and Budget, "is an evaluation of a body of scientific or technical knowledge that typically synthesizes multiple factual inputs, data, models, assumptions, and/or applies best professional judgement to bridge uncertainties in the available information."[80]

The Working Group was charged with

- Examining and analyzing the human factors in current policies, procedures, and practices within the field of forensic handwriting examination;
- Developing practices based on scientifically sound research to reduce the likelihood of errors in forensic document examination;
- Evaluating various approaches to quantifying measurement uncertainty within forensic document analysis; and
- Publishing findings and recommendations that include future research initiatives.

The Working Group met eight times over 2.5 years and heard presentations from experts in the areas of human factors; the weight of evidence in law, statistics, and forensic science; decision making and formulation of propositions; probabilities and likelihood ratios; and other relevant topics.

Working Group members were selected by NIST and NIJ staff in consultation with the Working Group co-chairs based on their expertise in the forensic sciences, understanding of human factors principles, background in handwriting examination and forensic document analysis practices and training, understanding of statistics in forensic science, and the use and acceptance of handwriting testimony in the courts. The Working Group consisted of an international group of forensic science experts in handwriting examination (working as sole practitioners or in larger forensic laboratories), legal scholars, forensic science academics, statisticians, cognitive scientists, and professional organization representatives.

Each chapter of this report was developed by a subcommittee and presented to the entire Working Group for review. The draft report was developed through a consensus process that allowed each Working Group member to comment on and influence all recommendations and

---

[79] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, U.S. Department of Commerce, National Institute of Standards and Technology (2012).

[80] Office of Management and Budget, *Final Information Quality Bulletin for Peer Review* (December 15 2004), 1, https://www.epa.gov/sites/production/files/2015-01/documents/omb_final_info_quality_bulletin_peer_review_2004_1.pdf.

text. The draft report was edited by a committee formed from a subset of the Working Group members and reviewed by a panel of independent experts not associated with the Working Group. The editorial committee then resolved all comments from the independent experts and presented the final draft to the Working Group for review and final consensus. The group, despite having differing viewpoints and diverse backgrounds, reached substantial agreement on many foundational issues not limited to the formal recommendations. Some topics discussed represent future directions and trends that may not be fully embraced by the entire group; particular chapters indicate these differences.

The Working Group focused exclusively on the analysis and comparison of handwriting, including cursive and hand-printed text, numerals, and signatures. The group did not address other aspects of questioned document examinations like analysis and comparison of ink and paper, typewritten text, and preprocessing techniques. The Working Group also did not consider graphology (the analysis of handwriting to infer a person's character), which is considered a pseudoscience.

In examining human factors, the Working Group considered trends likely to have a major impact on forensic document examination. The Working Group addressed the need for national training standards for FDEs and made recommendations for standardizing handwriting analysis report content and communicating report information to clients and courts. The Working Group also had robust discussions regarding the potential use and practicality of probabilistic interpretation (i.e., likelihood ratios) in the expression of handwriting opinions, because this method is employed in several countries globally.

A probabilistic interpretation of results or a determination that the evidence is inconclusive requires clear and careful explanations in both written reports and testimony; however, no consensus exists for how to define and express probabilities nor is there a single standard procedure for communicating such information. Although a probabilistic approach is more widely used outside the United States, the Working Group felt a discussion was warranted to assess whether this approach was appropriate and practical in the current setting as related to human factors considerations.

In surveying the human factors associated with forensic document examination, the Working Group acknowledged the shrinking and aging pool of FDEs. A 2017 survey of members of the American Society of Questioned Document Examiners (ASQDE) who were still active handwriting examiners at the time of the survey revealed that the average age of respondents was 57 years.[81] At the time of publication, the average age of all ASQDE members, including those retired but still contributing to the society, was 60 years with a median age of 63 years.[82] Under "professional, technical and scientific occupations," the median age is 42 to 44 according to data compiled by the U.S. Department of Labor.[83]

Across the country, forensic document examination units within crime laboratories are closing as demand shifts to other forensic disciplines like DNA analysis. The modern world's de-emphasis

[81] This survey was conducted by a Working Group member for the purpose of including in this report. There were 57 respondents. An earlier version of this report stated that the median age of examiners was 60 years.

[82] These data reflect all ASQDE member types, including life members, corresponding members, trainees, and provisional members ($N = 113$).

[83] "Labor Force Statistics from the Current Population Survey," updated February 8, 2017, https://www.bls.gov/cps/cpsaat18b.htm.

on handwritten communications continues to impact the field, as has the increasingly central role of automation both in aiding the FDE in analyzing handwriting and in capturing handwriting data, such as digital signatures. To adapt to these changes, FDEs may need to expand their expertise to other branches of forensic science, such as analyzing fingerprints and shoe and tire impressions, and they may need to gain more experience with automated systems.

Finally, the Working Group addressed fragmentation within the FDE community. Different FDE groups have strong differences in opinion about training requirements, partly because of their different modes of training. Some FDEs were trained in government or private laboratories, whereas others are self-trained or used distance learning. In the past, efforts have been made to establish a minimum training requirement[84] for all FDEs, but this training standard has not been universally accepted.

Some FDEs consider the minimum training standard as a guideline that does not apply to them, and others disavow any relevance of the standard to their work or have instead suggested their own standards. FDEs working in the private sector face an additional difficulty: balancing training requirements with the cost and time involved in meeting those requirements on a limited budget. As a result of these disparities, some FDEs have established their own professional organizations and certifying bodies, publish in separate journals, and rarely interact with other groups. The Forensic Specialties Accreditation Board (FSAB)[85] accredits the American Board of Forensic Document Examiners (ABFDE) and the Board of Forensic Document Examiners (BFDE). Other professional membership organizations that provide certifications, such as the National Association of Document Examiners and Scientific Association of Forensic Examiners, are not accredited by FSAB.

By including FDEs with widely different opinions on training requirements and those who work in a variety of settings (i.e., small private practices and large government laboratories) in its roster, the Working Group encouraged debate and dialogue between subject matter experts who had not previously had the opportunity to communicate with each other effectively. In doing so, the Working Group not only embraced diversity of opinion but forged a consensus on establishing best practices for training and other areas. This also enabled the Working Group to develop recommendations and suggested standards that can be universally applied to FDEs.

In addressing these concerns and making recommendations, this report is aimed at policy makers in federal, state, and local government, along with FDEs in private and public practice. Additionally, this report and its recommendations can be applied to international organizations.

The Working Group recognizes that many recommendations will take time to implement, and it is unreasonable to demand that laboratories of all types satisfy these recommendations overnight. Equally, it is unreasonable to expect that laboratories will suspend work and cease serving the legal community until and unless these recommendations are implemented. This report offers significant discussion on how recommendations can be implemented, including guidance to small and sole practitioner laboratories.

---

[84] *SWGDOC Standard for Minimum Training Requirements for Forensic Document Examiners,* Version 2013-1, (Scientific Working Group for Forensic Document Examination (SWGDOC), 2013).

[85] "Forensic Specialties Accreditation Board (FSAB) home page," updated April 5, 2020, 2019, http://thefsab.org/.

1. Members

The Working Group relied on contributions from many individuals to meet its charge. The opinions presented over the course of the Working Group's deliberation reflect personal experiences and views and do not express the official positions of the institutions with which members are affiliated.

Carolyne Bird, PhD, Science Leader, Document Examination, Forensic Science SA, Australia [Working Group Editorial Committee]

Brett M. Bishop, FDE, Washington State Patrol

Ted Burkes, FDE, Federal Bureau of Investigation (FBI) Laboratory [Chair of Working Group; Working Group Editorial Committee]

Michael P. Caligiuri, PhD, Emeritus Professor, University of California at San Diego; Department of Psychiatry [Working Group Editorial Committee]

Bryan Found, PhD, Chief Forensic Scientist, Victoria Police Forensic Services Department, Australia

Wesley P. Grose, Crime Laboratory Director, Los Angeles County Sheriff's Department [Working Group Editorial Committee]

Lauren R. Logan, Forensic Scientist II, Indiana State Police Laboratory [Working Group Editorial Committee]

Kenneth E. Melson, JD, Professorial Lecturer in Law, George Washington University Law School [Working Group Editorial Committee]

Mara L. Merlino, PhD, Associate Professor of Psychology and Sociology, and Coordinator, Master of Arts Program in Interdisciplinary Behavioral Science, Kentucky State University

Larry S. Miller, PhD, Professor and Chair, Department of Criminal Justice, East Tennessee State University

Linton Mohammed, PhD, FDE, Forensic Science Consultants, Inc., Burlington, CA

Jonathan Morris, Forensic Scientist, Scottish Police Authority Forensic Services, Scottish Crime Campus

John Paul Osborn, FDE, Osborn and Son, Middlesex, New Jersey

Nikola Osborne, PhD, Postdoctoral Scholar, Department of Criminology, Law and Society, University of California, Irvine [Working Group Editorial Committee]

Brent Ostrum, Senior FDE, Canada Border Services Agency

Christopher P. Saunders, PhD, Associate Professor of Statistics/Lead Signal Processing Engineer, South Dakota State University/MITRE

Scott A. Shappell, PhD, Professor and Chair, Department of Human Factors, Embry-Riddle Aeronautical University

H. David Sheets, PhD, Professor, Department of Physics, Canisius College

Sargur N. Srihari, PhD, Distinguished Professor, Department of Computer Science and Engineering, State University of New York at Buffalo

Reinoud D. Stoel, PhD, Netherlands Forensic Institute, the Netherlands [Working Group Editorial Committee]

Thomas W. Vastrick, FDE, Private Practice, Apopka, Florida [Working Group Editorial Committee]

Heather E. Waltke, **MFS, MPH**, Associate Director, OIFS, NIJ [Working Group Editorial Committee]

Emily J. Will, MA, FDE, Private Practice, Raleigh, North Carolina

Staff

Melissa Taylor, Study Director, Special Programs Office, NIST

Ron Cowen, Writer and Editor

Katherine Fuller, Desktop Publisher/Editing Specialist, Leidos

Christina Frank, Editor, Leidos

MacKenzie Robertson, Independent Consultant, Dakota Consulting, Inc.

Katherine Ritterhoff, MS, Project Manager, Leidos

2. About the Sponsors

NIJ is the research, development, and evaluation agency of the U.S. Department of Justice and is dedicated to researching crime control and justice issues. NIJ provides objective, independent, evidence-based knowledge and tools to meet the challenges of the nation's criminal justice community. NIJ's OIFS is the federal government's lead agency for forensic science research and development and administers programs that provide direct support to crime laboratories and law enforcement agencies. OIFS forensic science programs and initiatives provide resources for the creation of new, innovative, and emerging technologies through the integration of research and development, laboratory efficiency and capacity enhancement, and technology transition, which will increase the capacity of crime laboratories to process growing amounts of evidence effectively and expeditiously.

The NIST mission is to advance measurement science, standards, and technology. It accomplishes these actions for the forensic science community through its Special Programs Office's Forensic Science Program (FSP). The FSP directs research efforts to develop performance standards, measurement tools, operating procedures, guidelines, and reports that will advance the field of forensic science. The Special Programs Office also manages the Organization of Scientific Area Committees for Forensic Science (OSAC), which works to strengthen the nation's use of forensic science by facilitating development of technically sound

forensic science standards and promoting adoption of those standards by the forensic science community.

3. Organization of This Report

To understand better how human factors impact forensic document examination, the Working Group carefully annotated the process for conducting an examination and reporting the results. This process map, detailed in chapter 1, describes the current steps FDEs follow to reach a conclusion regarding a handwriting comparison or to determine that the evidence is insufficient to reach a conclusion. Throughout the remainder of this report, there will be additional discussions regarding the scientific foundations of handwriting examination, such as uniqueness, uncertainty, and repeatability, along with recommendations aimed at modifying the process map to reduce human error.

Meticulously comparing known and questioned documents, accurately interpreting the data, and understanding and correctly employing probability in reporting results are the fundamentals of a forensic document examination. Chapter 2 highlights how human factors can affect each component of the examination process and introduces the concept of bias in forensic analysis. Chapter 2 also discusses the currently available automated technologies to aid the FDE.

What are the tools and procedures FDEs should employ when writing a report about a questioned document? How can that report be most effectively communicated to the courts, whether through testimony or a written document? Chapter 3 addresses these questions, which may have significant consequences for reaching an accurate conclusion and conveying information so that it is interpreted correctly.

An effective quality assurance/quality control (QA/QC) program is critical for identifying, correcting, and preventing errors in forensic handwriting examinations. Chapter 4 outlines the requirements of a QA/QC program, including considerations for companies with only one or a few practitioners.

Education, training, and certification are basic tools to ensure the high quality and continued excellence of FDEs and to minimize the impact of human error on the examination process. Chapter 5 assesses the status of education, training, and certification, including recommendations to use these tools most effectively.

A good manager creates an environment in which errors can be acknowledged, identified, and corrected in an efficient, non-punitive manner. Chapter 6 focuses on the qualities that constitute an effective management system and discusses how managers can most effectively recognize and mitigate the negative impact of human factors.

Recommendations on the need for research appear in relevant chapters, and chapter 7 summarizes the recommendations made throughout this report.

4. Acknowledgements

Presenters and Discussants

The Working Group gratefully acknowledges the following individuals for their contributions to the development of this document through subject matter presentations or meeting participation.

These persons, however, were not asked to review or comment on the final report. Therefore, the views expressed in this report reflect those of the authors and not necessarily the views of those acknowledged here.

Peter Belcastro, MFS, Training Program Manager, Questioned Document Unit, FBI Laboratory

Dana M. Delger, JD, Staff Attorney, Strategic Litigation, Innocence Project

Itiel E. Dror, PhD, Principal Consultant and Researcher, Cognitive Consultants International

Lindsey Dyn, MFS, Document Examiner and Quality Assurance Program Manager, Questioned Document Unit, FBI Laboratory

Christina Frederick, PhD, Professor, PhD, and Master Program Coordinator, Human Factors and Systems, Embry-Riddle Aeronautical University

Cami Fuglsby, MS, Doctoral Student in the Department of Mathematics and Statistics, South Dakota State University

Melissa R. Gische, MFS, Physical Scientist/Forensic Examiner, Latent Print Operations Unit FBI Laboratory

Derek L. Hammond, MS, FDE, U.S. Army Criminal Investigation Laboratory

Hariharan Iyer, PhD, Mathematical Statistician, Statistical Engineering Division, Information Technology Laboratory, NIST

Steve Lund, PhD, Mathematical Statistician, Statistical Engineering Division, Information Technology Laboratory, NIST

Moshe Kam, PhD, Dean of Newark College of Engineering, New Jersey Institute of Technology

Jason Kring, PhD, Project Lead for Embry-Riddle's Mobile Extreme Environment Research Laboratory, Embry-Riddle Aeronautical University

Danica M. Ommen, PhD, Assistant Professor of Statistics, Iowa State University at Ames

Andrew Plotner, PhD, Visiting Scientist (28 May 2015 through 8 April 2016), FBI Laboratory, Counterterrorism and Forensic Science Research Unit

Honorable Ron Reinstein, JD, Former Arizona Superior Court Judge

William C. Thompson, PhD, Professor of Criminology, Law, and Society; Psychology and Social Behavior; and Law, University of California Irvine School of Social Ecology

Rigo Vargas, Questioned Documents Section Chief, Mississippi Crime Laboratory (Gulf Coast Regional Lab)

Nicholas Vercruysse, MS, Visiting Scientist (1 May 2014 through 11 August 2017), FBI Laboratory, Counterterrorism and Forensic Science Research Unit

Reviewers

This report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. Although the reviewers listed provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. Responsibility for the final content of this report rests entirely with the members of the Working Group.

John M. Butler, PhD, NIST Fellow, Special Assistant to the Director for Forensic Science, Special Programs Office

Edward J Imwinkelried, JD, Professor of Law Emeritus, University of California Davis School of Law

David H. Kaye, MS, JD, Distinguished Professor and Weiss Family Scholar, Penn State Dickinson School of Law

Michael Risinger, JD, John J. Gibbons Professor of Law, Associate Director, Last Resort Exoneration Project, Seton Hall University School of Law

Andrew Sulner, JD, MSFS, FDE, Forensic Document Examinations, LLC, New York

Mary Theofanos, MS, Computer Scientist, Material Measurement Laboratory, NIST

Pam Zilly, Crime Laboratory Director, Nebraska State Patrol—Crime Laboratory

This page intentionally left blank.

1. Handwriting Examination Process

Introduction and Scope

Forensic handwriting comparison, including the examination of cursive writing, hand printing, signatures, and numbers, is part of the broader field of forensic (or questioned) document examination. This forensic discipline draws on many types of expertise and scientific techniques. A document in this context is a tangible communication—a writing, drawing, or stamped impression on paper or another physical medium—and a questioned document is one whose authenticity, source of origin, or means of preparation is under investigation. The investigation can address the composition of paper, ink, or other materials. In addition, when the communication is handwritten, different aspects of the marks provide evidence about a document's potential writer. More specifically, an FDE may be called on to answer—or to supply information that would help a judge or jury answer—questions involving authenticity and writership,[86] like the following: Is the writer of the exemplars also the writer of the questioned document(s)? Were the questioned documents written by only one individual?

A handwriting examination involves human perceptions and interpretation of the similarities and differences among the questioned writing and the standards or exemplars from known individuals. Using a process map (figure 1.1) as a description of the current practice, this chapter describes how an FDE conducts handwriting comparisons. The map is presented to aid discussion about key decision points in the procedure.

The Working Group believes that some of the process map steps can and should be modified or informed by data to reduce the adverse effects of human factors on work product quality. The Working Group's recommendations in this regard appear throughout the other chapters of this report, section 2.3 discusses an alternate evaluation approach.

1.1. The Conventional Process of Forensic Handwriting Comparison

The early pioneers of forensic document examination, such as Albert S. Osborn, were skilled penmen who worked at a time when handwriting was taught as a necessary business skill. They could tell when writers deviated from the various copybook systems being taught. They referred to the features contained within copybook styles as class characteristics and the deviations from the copybook style as individual characteristics. Their system of handwriting identification was based on ascertaining the individual characteristics and determining whether they were indicative of one writer or two or whether there had been an attempt to simulate another person's handwriting characteristics.

Over time, however, the priority teaching of handwriting as a skill diminished, the number of copybook systems taught in schools has increased, and people who were taught different copybook styles are more geographically dispersed. As a result, a more contemporary view is that determining the particular copybook style learned by an unknown writer would be extremely

[86] Please refer to the note on page XVI.

difficult, if not impossible.[87] This position is further supported by research on the variety of handwriting systems currently being taught in Canada.[88]

Despite the perceived difficulty in determining copybook styles, the conventional belief in individuality persists among FDEs—that is, the assumption that no two writers share the same combination of handwriting characteristics[89] and that before reaching adulthood, people will establish a consistent writing habit.[90] New theories based on the neurobiological principles underlying handwriting variation that have emerged within the last 2 decades explain the handwriting process further (see section 2.3).[91]

The conventional process for answering questions about writership involves perceiving and measuring selected features in the handwriting specimens, ascertaining how these features differ across specimens, and interpreting the significance of the similarities and differences. Although some aspects of handwriting examinations may involve physical measurements, FDEs more often rely on relative measurements—the estimation of features proportionally to one another. Relative measurements can include size, spacing, and the slant of features. The FDE's comparison and evaluation of the writing may result in an opinion ranging from eliminating a given individual as the writer of questioned writing to positively identifying the individual. Although the Working Group is necessarily critical of some aspects of the conventional process (see chapter 3), it is presented here as the starting point from which to develop recommendations to improve the discipline.

1.2. The Process

During an examination, an FDE reaches their opinion through a process that involves many steps, shown in the process map (figure 1.1). The Working Group developed the process map in collaboration with others in the FDE community to represent current practices in the United States. The steps outlined are typical of a routine handwriting examination case and are presented in a linear fashion; however, in practice, the sequence of steps may vary, and several steps or examinations may be conducted in parallel, and additional steps may be necessary in some cases.

Other methods used in handwriting examination are described in a modular approach developed by the Document Examination Specialist Advisory Group of Australia and New Zealand[92] and are documented within the *Best Practice Manual for the Forensic Examination of Handwriting*

[87] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 27.

[88] Lindsay Holmes, "Handwriting Instruction in Canadian Schools as Prescibed by Provincial and Territorial Ministries of Education," *Canadian Society of Forensic Science Journal* 43, no. 1 (2010): 9–15, https://doi.org/10.1080/00085030.2010.10757616.

[89] D. Harrison, T. M. Burkes, and D. P. Seiger, "Handwriting Examination: Meeting the Challenges of Science and the Law," *FBI Forensic Science Communications* 11, no. 4 (2009), https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2009/review/2009_10_review02.htm.

[90] H. Sieden and F. Norwitch, "Questioned Documents," in *Forensic Science: An Introduction to Scientific and Investigative Techniques*, ed. S. H. James, J. J. Norby, and S. Bell (Boca Raton: CRC Press, 2014), 451.

[91] B. Found and D. Rogers, "Contemporary Issues in Forensic Handwriting Examination: A Discussion of Key Issues in the Wake of the Starzecpyzel Decision," *Journal of Forensic Document Examination* 8 (1995): 483–92.

[92] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83.

produced by the European Network of Forensic Science Institutes (ENFSI).[93] However, the general procedure for all approaches includes

- Analyzing the features of the questioned writing and known standards both macroscopically and microscopically;
- Noting conspicuous features like size, slant, and letter construction, and more subtle characteristics like pen direction, the nature of connections between letters, and spacing between letters, words, and lines;
- Comparing the observed features to determine similarities and dissimilarities; and
- Considering the degree of similarity or dissimilarity and the nature of the writing (quality, amount, and complexity), evaluating the evidence, and arriving at an opinion regarding the writership of the questioned writing.

[93] European Network of Forensic Science Institutes (ENFSI), *Best Practice Manual for the Forensic Examination of Handwriting* (November 2015), https://enfsi.eu/wp-content/uploads/2016/09/2._forensic_examination_of_handwriting_0.pdf.

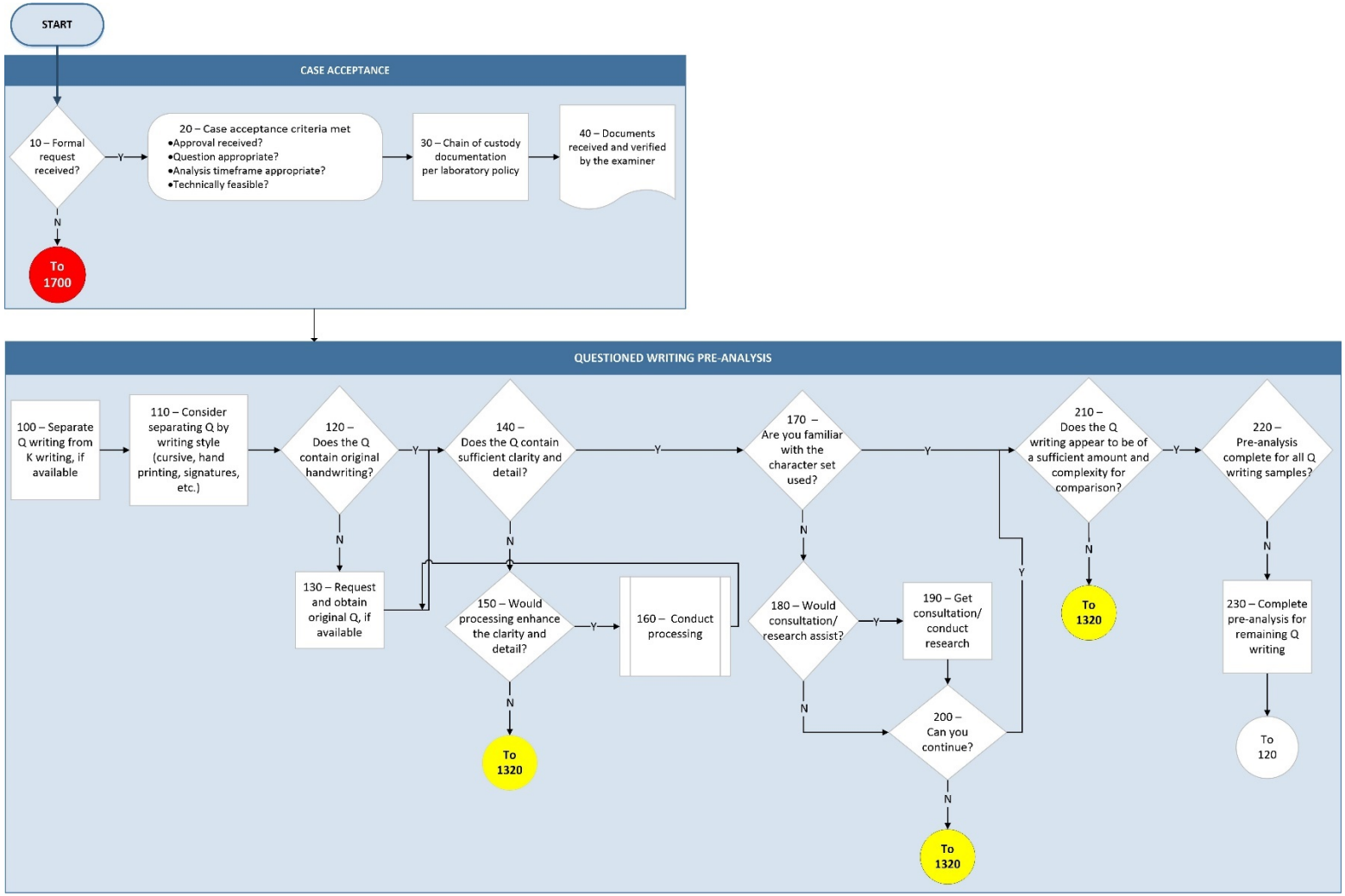


Figure 1.1: Handwriting examination process map

This diagram documents the steps of the examination process as currently practiced by the U.S. handwriting examination community. The numbers in each box correspond to steps that are more fully described throughout the report. The purpose of this process map is to facilitate discussion about key decision points in the handwriting examination process. (Map continued on next page.)

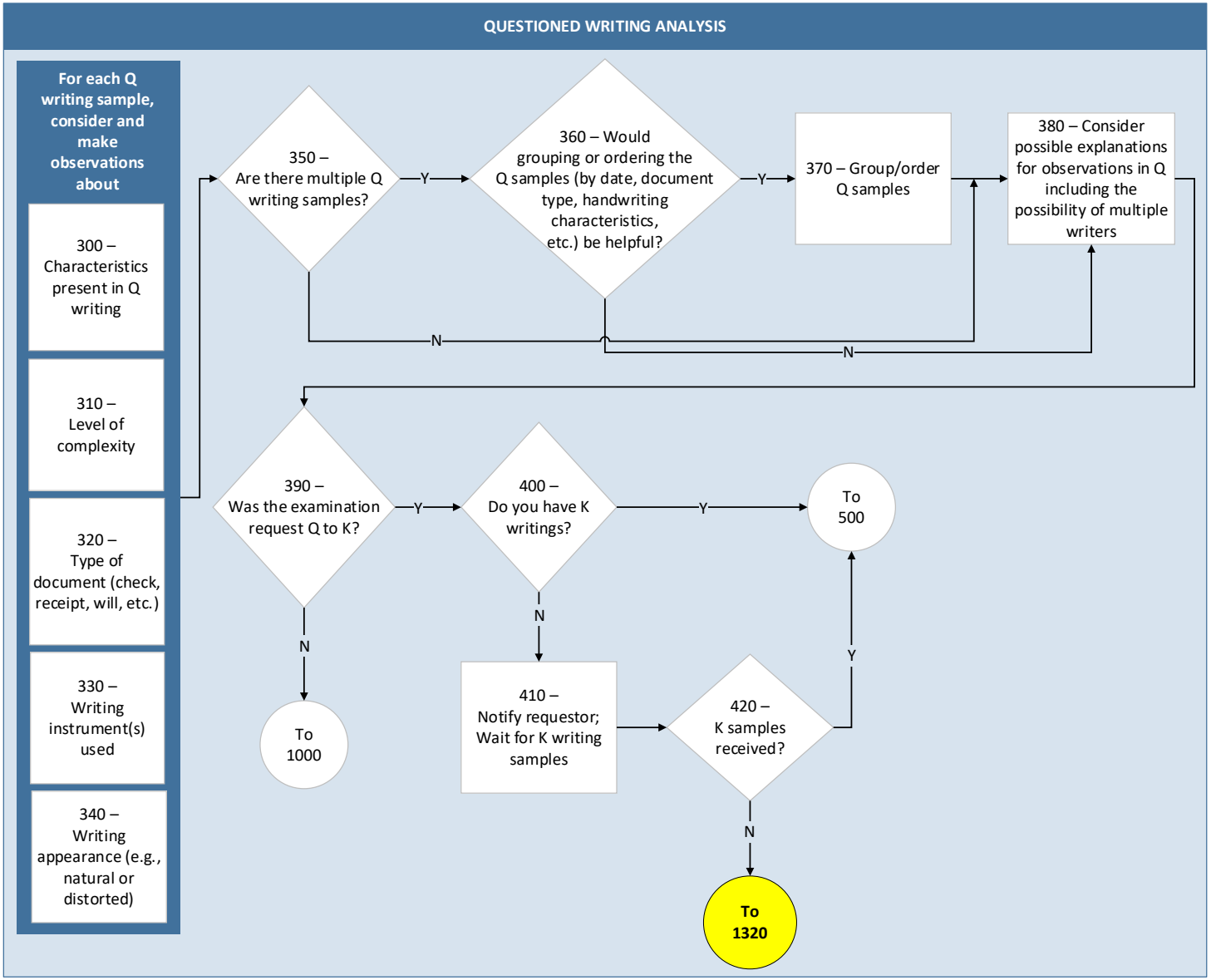


Figure 1.1: Handwriting examination process map (Continued)

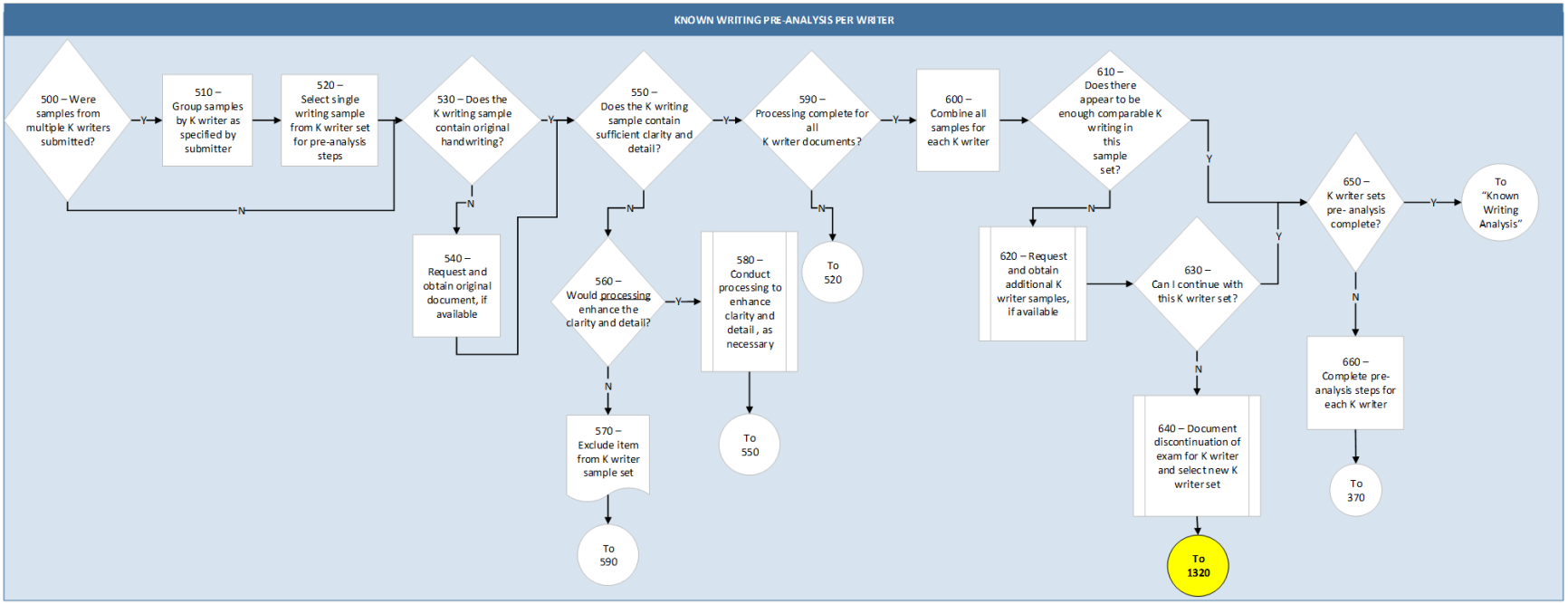


Figure 1.1: Handwriting examination process map (Continued)

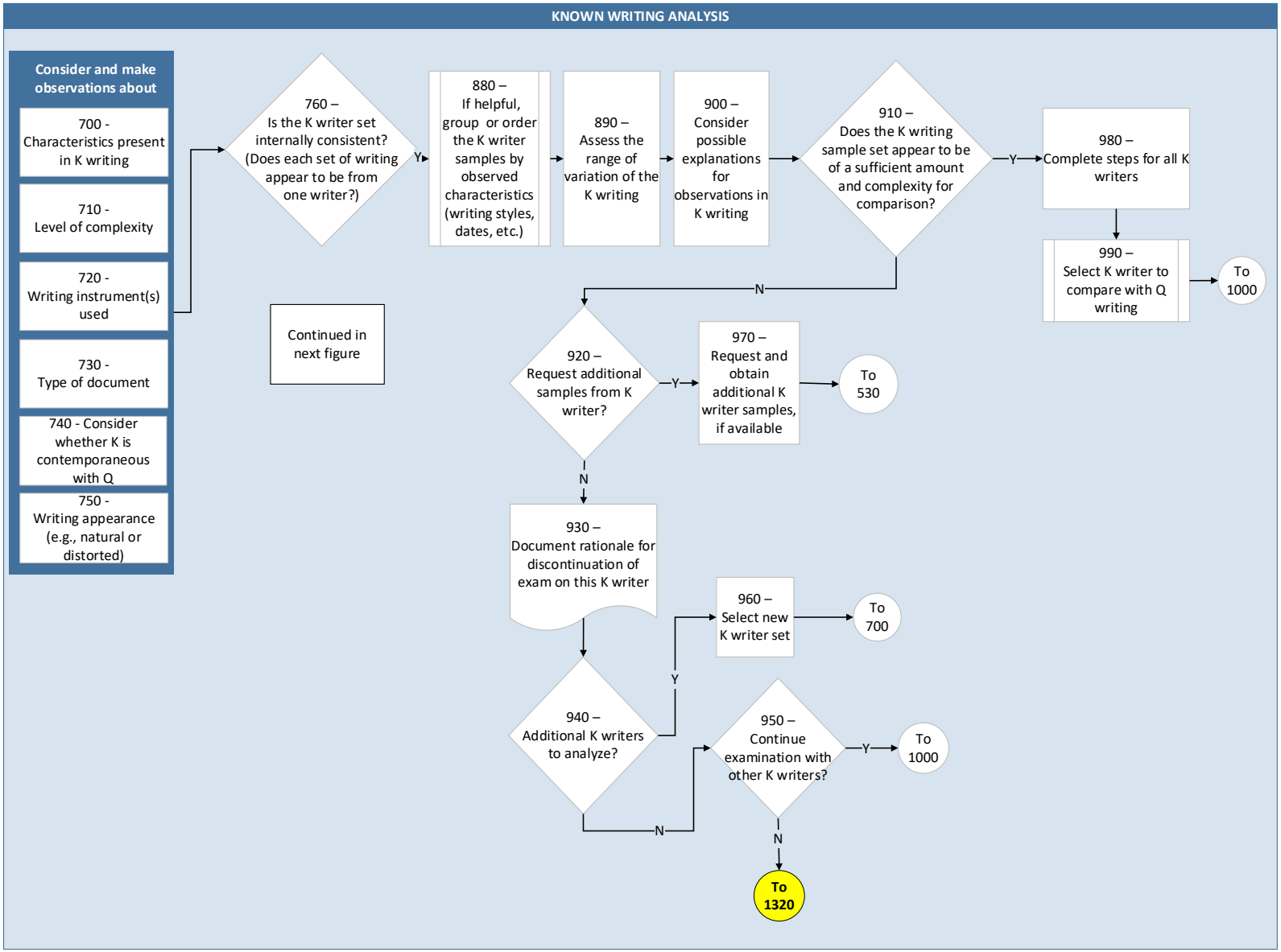


Figure 1.1: Handwriting examination process map (Continued)

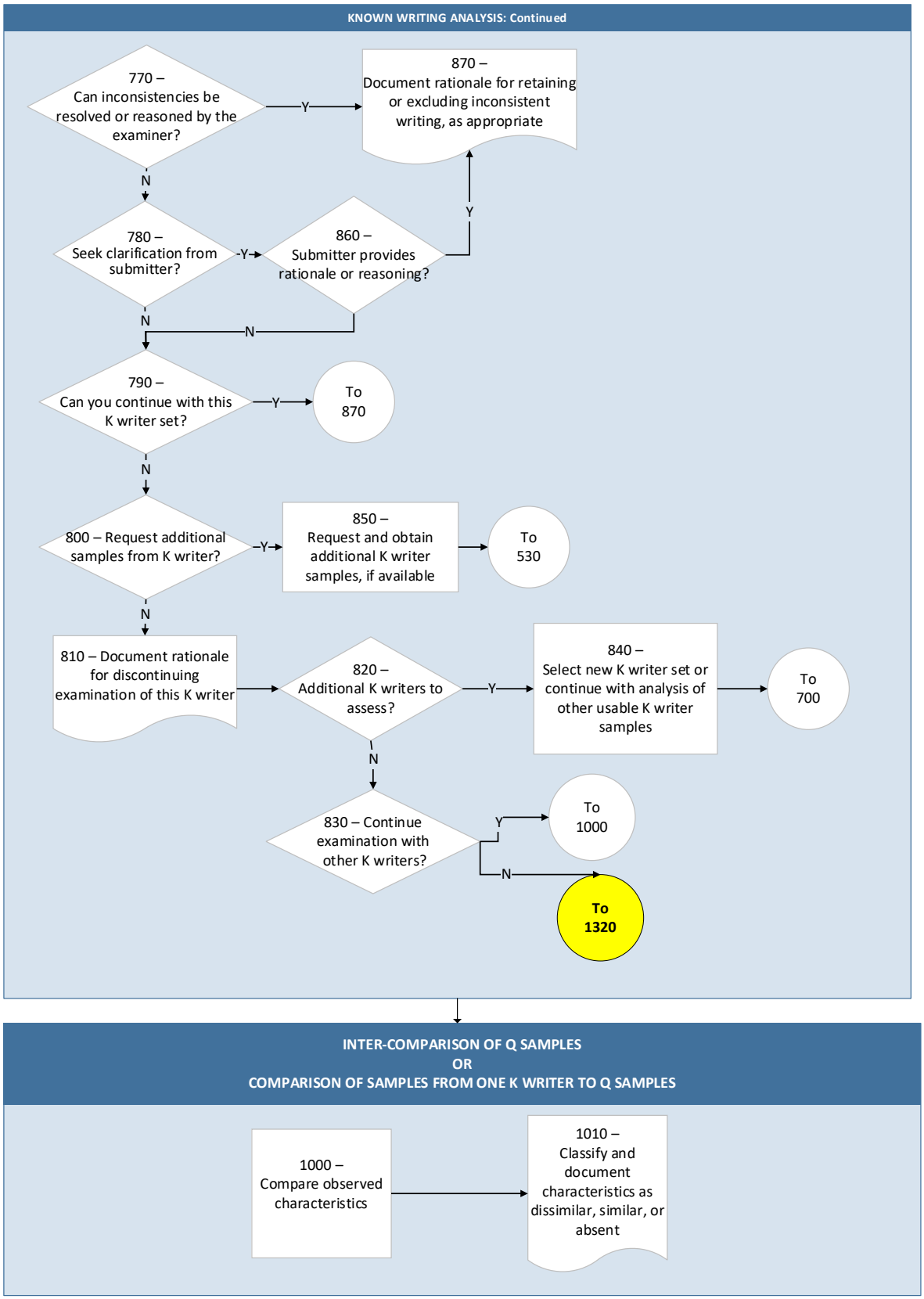


Figure 1.1: Handwriting examination process map (Continued)

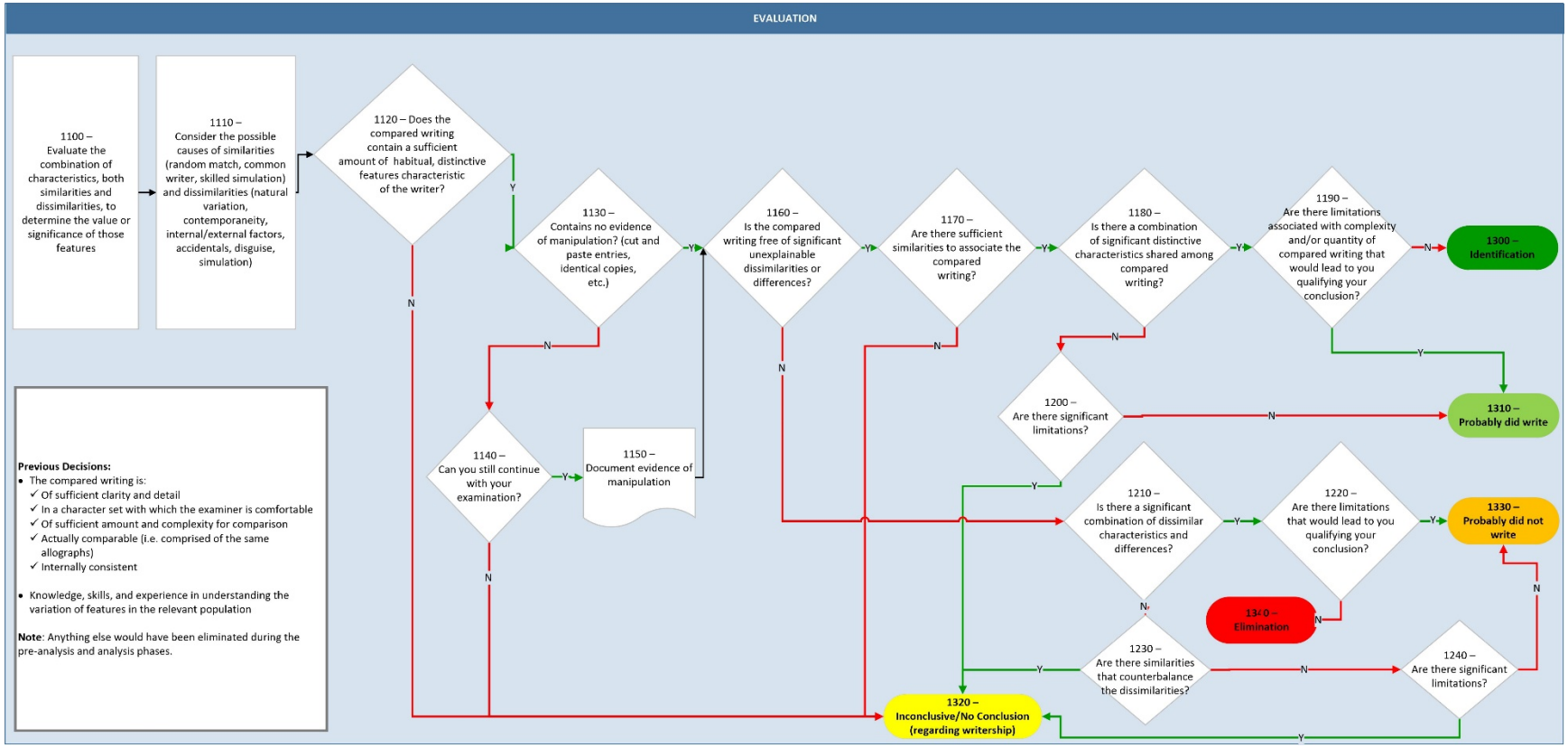


Figure 1.1: Handwriting examination process map (Continued)

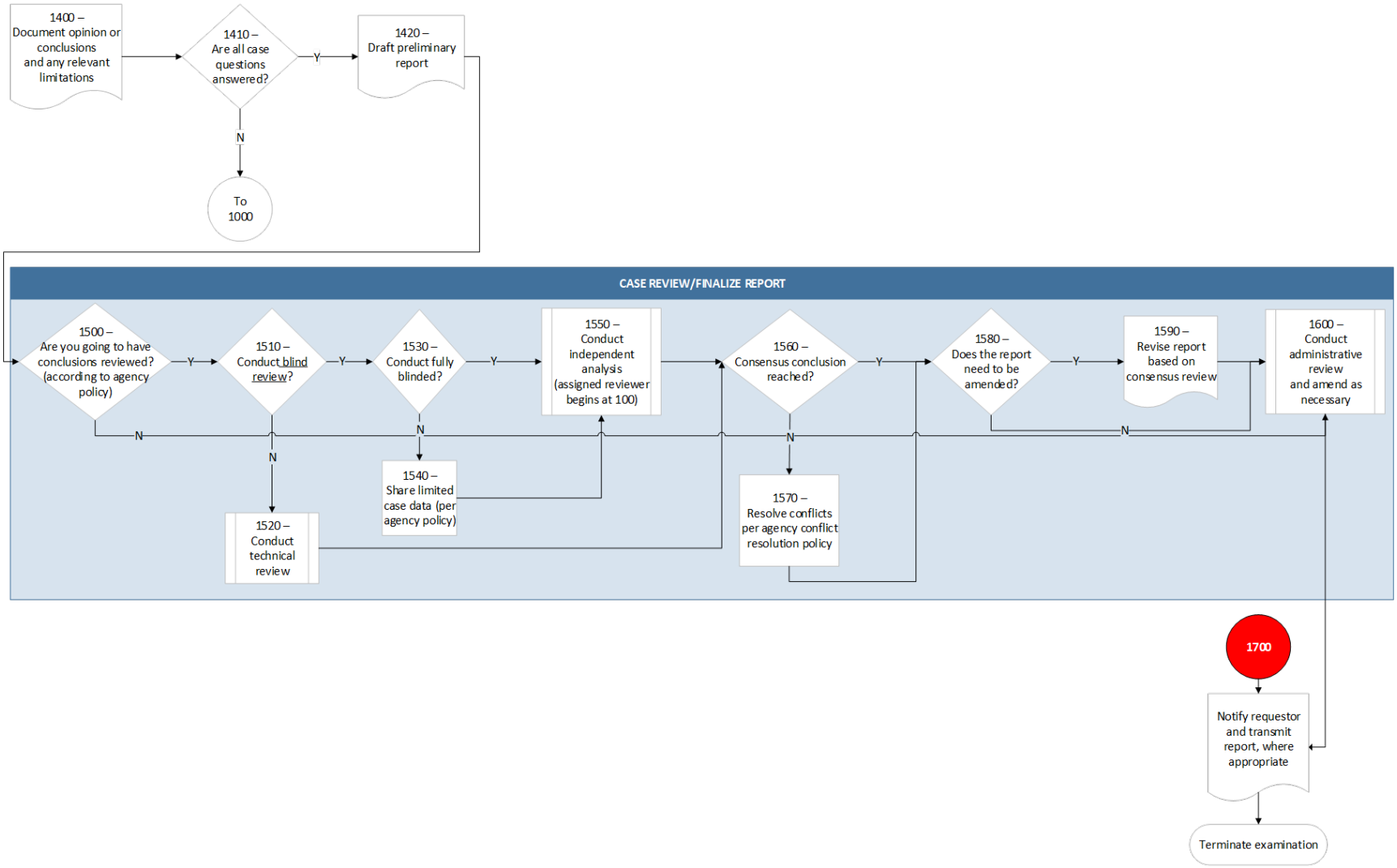


Figure 1.1: Handwriting examination process map (Continued)

1.2.1. Case Acceptance [Steps 10–40]

Documents are submitted to a laboratory for examination along with a formal request outlining the question to be answered. The acceptance procedure for the documents depends on the laboratory. Larger laboratories may have a central evidence receipt unit in which an FDE (who may be either a forensic document examiner or a forensic handwriting examiner) reviews the documents. The FDE decides whether the documents are properly packaged and labeled to establish a chain of custody. The evidence undergoes a triage process to determine the order of examinations (e.g., handwriting, latent prints, and DNA).

Latent print and DNA processing may interfere with, or render impossible, examinations like indented impressions or ink comparisons. Therefore, depending on the case circumstances and required examinations, crime laboratories may choose to send the documents to the FDE first. In these cases, appropriate precautions are taken to prevent evidence contamination with respect to the other examinations. In a smaller laboratory, the FDE may receive the documents and conduct an initial review of the material. If the documents are suitable for examination, the FDE accepts the documents, assigns a case number, and records the submission. If unsuitable, the FDE rejects the case (giving a reason) or discusses ways to improve the submitted material (e.g., by requesting the addition of handwriting exemplars) and records the request where appropriate.

At the time of submission, the laboratory or FDE decides whether the timeframe requested for the examination is feasible. If not, the case is rejected, or a suitable timeframe is negotiated. For urgent cases or where life or liberty are a factor (e.g., kidnappings or terrorist threats), the laboratory may expedite the examination process. FDEs may expedite urgent civil cases by giving their clients advice or verbal opinions.

After the documents are received, they are labeled with specific designations (i.e., questioned and known). The method of identifying the document, such as marking directly on the document or on copies of the documents, is determined by the laboratory's policy. The FDE should itemize and note the condition of all documents received.

FDEs usually work with two sets of documents: the questioned (sometimes referred to as Q) documents to be evaluated and the known (sometimes referred to as K) documents produced or acquired for the purpose of comparison. For cases where no known writing is available, an inter-comparison of the questioned documents may be possible to determine if they were written by the same individual. The process map provides a pathway for both types of comparison.

1.2.2. Questioned Writing Pre-Analysis [Steps 100–230]

The questioned documents are separated from the known documents, if available. In some cases, only questioned documents will be submitted. For example, if a serial bank robbery case has no suspect, and the investigator wants to know if all the demand notes were written by one person.

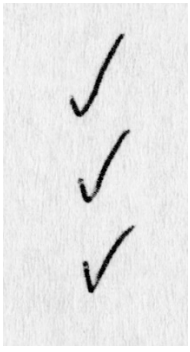
The FDE reviews the questioned documents and sorts them by handwriting type (e.g., signatures, cursive, or hand printing). The FDE also determines if the questioned documents are originals or copies; if copies, the FDE requests the originals from the submitter. In cases where the originals are only available at the document custodian's location, such as in court or an attorney's office, the FDE may conduct an off-site examination.

Regardless of whether original or copies of documents are available, the FDE determines if the available questioned documents are of adequate quality for a meaningful examination. Limitations in the amount or quality of the questioned documents generally cannot be improved upon, with the exception of enhancement of visibility of the line trace (e.g., image processing scans of faded entries).[94] If the questioned document quality is inadequate and enhancement provides insufficient improvement, then the FDE stops the examination and reports "no conclusion," with the reason (i.e., insufficiency of the questioned material) clearly stated. Ideally, this conclusion should be drawn before the known writing has been seen and with no knowledge of the context of the case (rationale outlined in section 2.1.3).

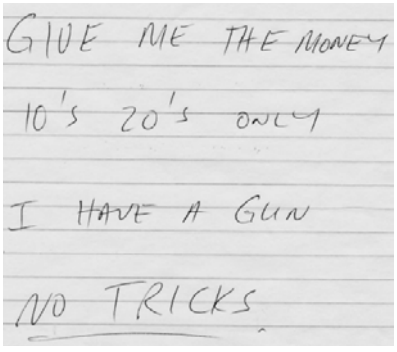
If the questioned documents are of adequate quality or enhancement improves the quality to a useful level, the FDE then determines their familiarity with the character set. For example, an English-speaking FDE who does not read any other languages will probably not be sufficiently familiar with Arabic script or Chinese characters to undertake a meaningful handwriting comparison of these. However, the FDE may consult resource documents or other FDEs to determine if the examination can proceed. If consultation and research do not help, then the FDE discontinues the examination and gives a "no conclusion" report, clearly stating the reason for being unable to continue with the examination.

If provided questioned material that is clearly visible and in a familiar character set, the FDE assesses whether the handwritten material has the quantity and complexity needed for an examination. For example, a questioned document that has a few generic check marks (as illustrated in figure 1.2A) may lack the quantity and complexity required for an examination. The document depicted in figure 1.2B, however, has an adequate amount of complex handwriting for examination.

Pre-analysis is repeated for each questioned document. At the end of this stage of the process, the FDE may have one or more questioned documents suitable to analyze in detail.



A



B

Figure 1.2: Generic check marks considered too simplistic for a meaningful examination (A) and more complex handwriting suitable for an examination to proceed (B).

[94] *SWGDOC Standard for Use of Image Capture and Storage Technology in Forensic Document Examination,* Version 2013-1, (Scientific Working Group for Forensic Document Examination (SWGDOC), 2013). Section 7.9.5.

1.2.3. Questioned Writing Analysis [Steps 300–420]

In the analysis phase for questioned handwriting samples, the FDE analyzes each questioned document separately. The FDE observes and notes characteristics of the handwriting as described in table 1.1 (and defined in the Glossary) and any relationships between them. These relationships include the letter formation, letter size, and inter-word and intra-word spacing, which affects the lateral expansion or horizontal dimension of words. A fundamental belief among FDEs is that these features are more variable across the writing of different individuals than within repeated writings of the same individual, but the statistical properties of these variable features have not been rigorously studied.[95] Section 2.3.1 discusses feature selection, and section 4.2.7 outlines the importance of documentation.

Table 1.1: Handwriting characteristics routinely considered during a handwriting examination[96]

Characteristics of handwriting style[97]	Characteristics of execution
• Arrangement or layout on the page • Connecting strokes • Construction • Design • Dimensions (including proportions) • Slant or slope • Spacing • Class • Allographs	• Abbreviations of words • Alignment • Commencements and terminations • Diacritics and punctuation • Embellishments • Line continuity • Line quality (smooth and fluent to tremulous and erratic) • Pen control (including pen hold, pen position, and pen pressure) • Complexity • Writing movement (including angularity) • Stroke order • Legibility or writing quality (including letter shapes or forms)

The FDE then determines the range of variation in handwriting characteristics seen in each questioned handwriting sample. The range is the extent to which the habits of the writer are either reproduced or vary on repeated occasions, which can affect all of the characteristics in table 1.1, from the construction of letters and numbers to slant, alignment, and line quality. For

[95] A preliminary study is reported in M. E. Johnson et al., "Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics," *Journal of Forensic Sciences* 62, no. 1 (January 2017), https://doi.org/10.1111/1556-4029.13248, https://www.ncbi.nlm.nih.gov/pubmed/27864959.

[96] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 136–38.

[97] With the possible exception of construction, these are the aspects of writing that play a significant role in the overall pictorial appearance of handwriting. Differences in construction do not necessarily alter the overall appearance.

example, figure 1.3 illustrates six forms of the letter "E" with different basic constructions. Using one or two of these forms is an example of narrow variation. Using three or four is considered a wide range of variation and using five or six of the forms would not be expected in one writer's habit (in the absence of deliberate change).

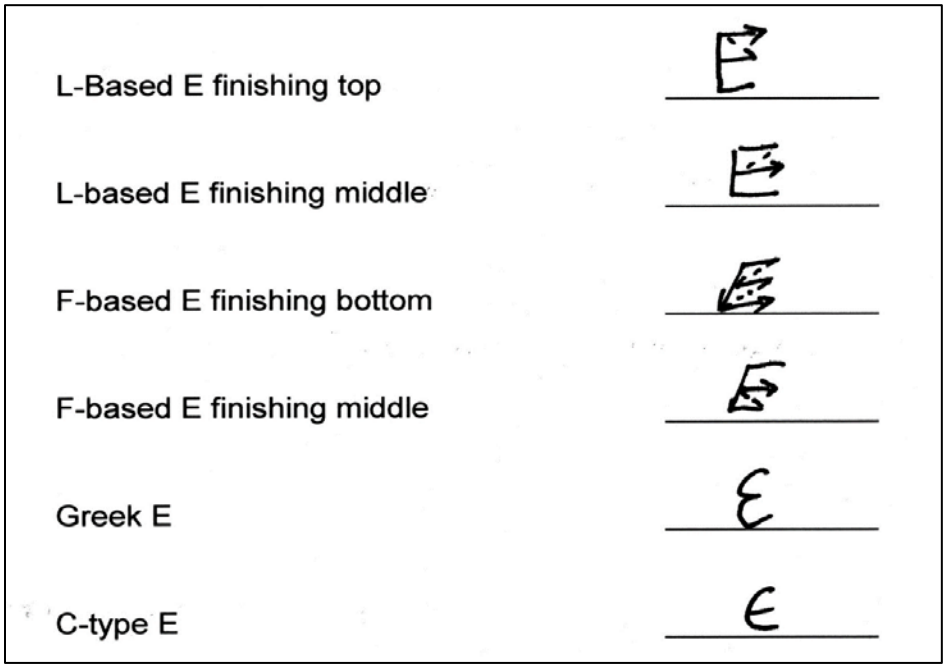


Figure 1.3: Differences in construction of the uppercase letter "E"

Figure 1.4 shows one example of what can be considered a normal, natural range of intra-writer variation in the uppercase letter "E."

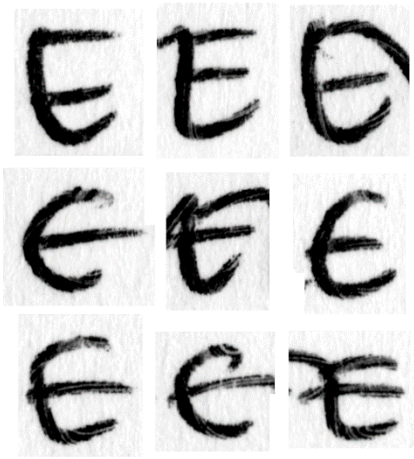


Figure 1.4: A range of natural variation in one writer's uppercase letter "E"

During the analysis, the -+FDE notes the frequency of occurrence, or persistence, of a given habit. For example, the position of a letter within a word might determine the use of a particular allograph.

The FDE also considers two other characteristics of the writing sample, rather than the writing itself: the type of document (e.g., letter, check, will) and the writing instrument(s) used, as these may affect the appearance of certain handwriting characteristics. The FDE also looks for evidence of distortion and considers possible explanations like the influence of alcohol or drugs/medication, unnatural writing positions, or disguise. If distortion appears to be present, the FDE will note it and should then determine whether it is possible to establish if the distorted writing is or is not natural writing. If the writing is not natural (or if it is impossible to establish whether the apparently distorted writing is natural writing), the FDE determines whether it is suitable for comparison. If the available questioned

writing is not suitable for comparison to known specimens, the FDE reports this as inconclusive/no conclusion (step 1320 of the process map).

After observing the characteristics of each questioned sample, the FDE assesses the range of variation displayed in a single questioned document or among many questioned documents to ensure that it falls within the expected range for a single writer, under the relevant conditions defined in the requested examination. If the range of variation exceeds what the FDE expects for a single writer, the questioned documents may then be sorted further into groups based on handwriting characteristics. The objective is to determine whether sets of writings share common handwriting features. Within each resulting group, the FDE ascertains the nature of the features and their range of variation in the writing.

The questioned writing samples may also be ordered or grouped based on date, document type, or another parameter the FDE deems useful.

During the analysis, the FDE should provide a written record that supports the conclusions with regard to the questioned documents. In particular, if the documents are suitable for comparison to known writings, the basis for this conclusion should be revealed by indicating which features the FDE believes will be useful in the later comparison phase of the process. This could be accomplished, as it is for latent fingerprints in some laboratories, by marking features to be compared on a photocopy of the questioned sample. This, however, does not prevent the use of additional features identified during the comparison phase.

1.2.4. Known Writing Pre-Analysis [Steps 500–660]

Known handwriting samples can either be requested (prepared specifically for comparison) or collected (normal daily writing). Each has advantages and disadvantages. Requested exemplars obtained for the matter at hand can be tailored to exhibit the same format, style, letters, letter combinations, word forms, and sentence structures as the questioned handwriting. In some cases, submitting parties have subjects complete pro forma exemplar documents. These are pre-set documents that contain instructions on what to write[98] and in what format. For example, the subject may be instructed to complete the exemplar in uppercase letters only. The exemplar documents are designed to capture many handwriting characters and their combinations. These documents usually supplement case-specific exemplars, but they can be used as a substitute if the case submitter does not want the subject to know the content of the questioned document.

The acquisition of requested samples generally proceeds in the following manner: (1) allow the subject to sit comfortably, (2) allow the subject to replicate the original (questioned) writing position (if known), (3) avoid having the subject see the questioned writing, (4) provide writing instruments[99] and materials[100] similar to those used to produce the questioned handwriting, and (5) have the subject produce multiple documents similar in format, style, and content to the

[98] Some examples of standard texts for request writings are given in Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 253–55.

[99] Most exemplars are generated using ballpoint pens. If the questioned writing was generated using a less common writing implement (e.g., a pencil or crayon), the subject should be requested to repeat the writings using this type of device.

[100] For example, if the questioned writing is text on a lined page, similar lined pages should be used.

questioned document(s).[101] The handwriting sample text can be dictated or provided in written/printed form. As the subject completes each page of exemplar writing, the individual collecting the handwriting signs and dates the document and removes it from view. FDEs are not generally responsible for acquiring known samples or verifying that the material submitted comes from the known individual.

Requested exemplars, either tailored or pro forma, are unlikely to exhibit the full range of natural writing because they are usually executed in a single sitting. Moreover, they may be atypical because of the attention placed on the writing act, the potential stress of the situation, and the opportunity for the writer to disguise their normal writing habits. For these reasons, collected writing is often preferable.

Collected exemplars, also known as normal course-of-business writings, are made during day-to-day activities. They are unlikely to be the product of disguise (particularly those collected before the time that a questioned sample of handwriting was purportedly written), and an ample collection is likely to show the full range of normal variation. In comparing collected exemplars to questioned handwriting, the style of writing is important. In general, signatures should only be compared with signatures, uppercase with uppercase, cursive with cursive, and printed writing with printed writing. As such, collected samples must include writing in the same format and style as the questioned material.

Other considerations that affect the value of collected exemplars might include the writing surface, writing instrument, and the purposes for which they were generated. It is useful for the collected exemplars to represent normal writing activity both before and after (and close to) the date(s) of the questioned writing(s). Collected handwritten text and signatures come from many sources.[102]

The pre-analysis procedure for known documents is analogous to that for the questioned documents, with the added first step of grouping the samples by known writer (if there is more than one) as specified by the case submitter.

The FDE proceeds through the pre-analysis procedure for each known writer individually. Like the questioned writing pre-analysis, the important questions asked are

- Do the known writing samples contain original handwriting? and
- Does the known writing contain sufficient clarity and detail for an examination to proceed?

In addition, the FDE determines if there are enough comparable known materials (for each writer set) to proceed with an examination. Primarily, comparability relates to the handwriting style or design (e.g., uppercase and lowercase hand printing, cursive) but also encompasses the

[101] For example, if the questioned writing is a signature of the subject's name, then the subject will be asked to provide several signatures (one per page). If the questioned writing is uppercase handwritten text, then the subject will be asked to write specific content in uppercase letters.

[102] For example, address forms, affidavits, business agreements, credit and insurance applications, charge account forms, membership applications, passport applications, work and school assignments, attendance records, banking documents, general business correspondence, recipes, credit card documents, grocery lists, guest registers, hospital records, identification cards, leases, mortgages, personnel records, greeting cards, post cards, tax returns, time sheets, and wills

characters (letters, numbers, and symbols or signs) present, the relative time between the writing of the questioned and known samples, and the form of the document(s) (see figures 1.5 and 1.6).

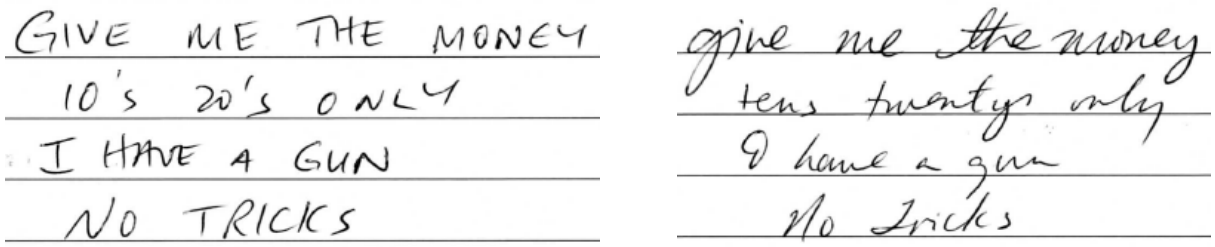


Figure 1.5: Handwritten entries that are not comparable even though they contain the same letters because they do not contain the same allographic form of letters

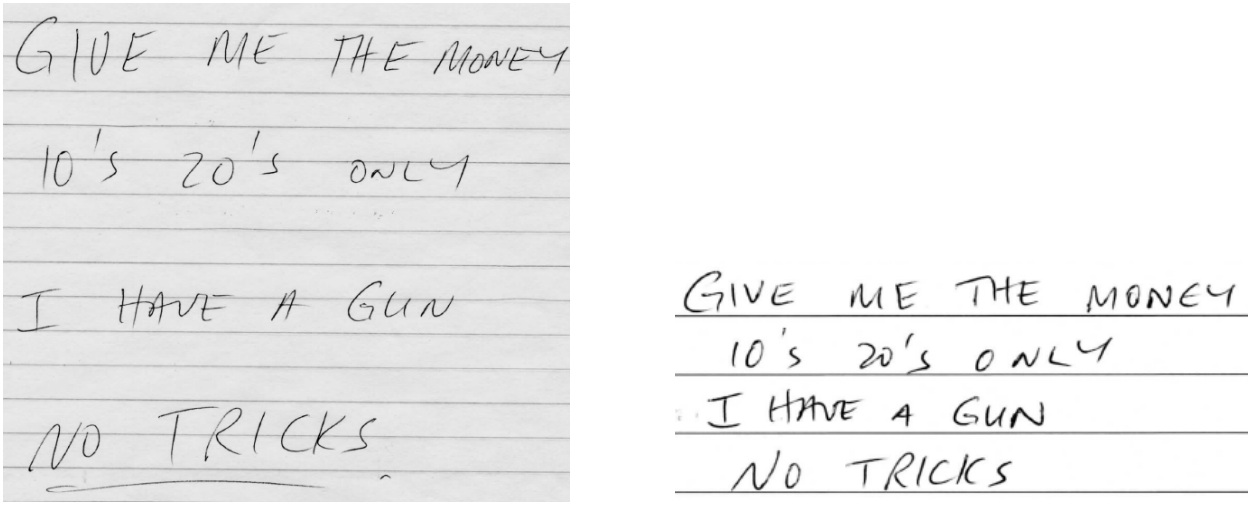


Figure 1.6: Handwritten entries that are comparable because they contain the same allographic form of letters; both are written in uppercase hand printing with the same letters and numbers present

Known samples of an individual must be of sufficient[103] quantity and quality to enable the FDE to compare them with questioned samples. If they are limited such that they do not capture natural variation or contain appropriate features for a comparison to be undertaken, the FDE may ask the submitting party for more known documents from the writer. Even if enough specimens are provided, the FDE may deem them as inadequate for comparison if they are not contemporaneous with the questioned writing. For example, if the questioned writing was written in 2017 and exhibits poor line quality, possibly because of age and illness, specimens from 20 years ago may not represent the writer's handwriting characteristics and range of variation in 2017.

Whether a known sample is wholly appropriate for comparison is difficult to determine objectively, may depend on the specific case, and involves the FDE's personal judgement. In an ideal setting, the conditions for selecting the reference material would be clearly defined in

[103] FDEs subjectively determine sufficiency, without reference to explicit criteria, as these do not currently exist.

advance. In practice, there are no generally accepted standard procedures. For example, the minimum number of known signatures recommended in the literature[104] ranges from 6 to 20 and for extended writing, a minimum of one to six pages. Generally, the FDE will prefer to see as many known specimens as are available.

If a known writer set does not contain enough clear, comparable writing to continue with the examination, the FDE discontinues the process for this known writer and reports the reason(s) why.

If the FDE determines that an examination can proceed, then the steps for analysis of the known writing are followed.

1.2.5. Known Writing Analysis [Steps 700–990]

A key first stage of the known writing analysis is to screen the exemplar writings of one individual for internal consistency or for possible writings from multiple individuals. This is an intra-comparison of the known documents for each known writer set. Quite often, documents submitted as bearing the known handwriting of one writer actually contain writings of multiple individuals. A typical example of this is a phone or address book. Unusual variations or inconsistencies in the exemplars may prompt an FDE to question the case submitter about the veracity of the samples, which may lead to exclusion of certain known writings or a request for more exemplars from specific known writers.[105] In some cases, the submitter may not provide clarification, and the FDE may not be able to continue with the known writer set. If additional exemplars for the specific known writer are not available, the FDE should document the rationale for discontinuing examination of this known writer. If clarification of the inconsistencies in the exemplars has not been obtained but the FDE can continue with the known writer set, then the FDE divides the writing samples from within the known writer set into groups based on handwriting features potentially belonging to different writers. The FDE should document this grouping and the rationale for continuing with the examination in this way. Again, additional grouping of samples by date, type, or handwriting style may be useful at the analysis stage of the process.

Just as for questioned writing analysis, the FDE should observe and note handwriting characteristics of each known writer to determine the nature and range of variation in these features. Once the FDE has (what is believed to be) an adequately representative sample set written by one writer, they then determine whether the sample is of sufficient amount and complexity for comparison. If so, the FDE proceeds with the known writer set to the next stage of the process along with the questioned writing sample(s).

---

[104] D. Ellen, *Scientific Examination of Documents: Methods and Techniques*, 3rd ed. (Boca Raton: CRC Press—Taylor & Francis Group, 2006), 83., "[The subject] should be asked to write the required passage at least five or ten times." Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 247., "For skilled or practised hands, a half dozen signatures or one or two pages of extended writing might prove adequate." Kelly and Lindblom, *Scientific Examination of Questioned Documents*, 136., "Therefore, if we are to ensure that the request specimens portray the natural handwriting variation of the individual . . . it is necessary to have the writer furnish at least five or six pages of continuous handwriting or 20 or more signatures"

[105] However, removing apparent outliers without further justification could bias subsequent comparisons toward a conclusion that the questioned handwriting is not authentic.

1.2.6. Comparison of Questioned and Known Samples [Steps 1000–1010]

Although the comparison stage of the process can be between two or more questioned writing samples or between questioned and known writing samples, the language used in the following description will assume that the FDE has both questioned and known samples. The process is the same for both scenarios.

If a case has multiple known writers of interest, the FDE can employ various methods for selecting the order of known writer sets for comparison against the questioned writing sample(s). Some FDEs take the known writers in either a random order or in order by the exhibit number or some other factor unrelated to the features being compared. Other FDEs select the known writer set that displays the most similar features to the questioned writing based on a preliminary assessment and begin the comparison and evaluation process with that "best match" set. Thus, the ordering of comparisons in a multi-known writer case may be influenced by human factors. In routine casework, these later stages of the process will be repeated for each known writer set.

The FDE then compares the characteristics of the questioned writing and the selected known writing using side-by-side comparison or by referencing a predefined set of features. The FDE looks for and documents feature similarities and dissimilarities and absent characters (i.e., characters present in one but not both samples, or absent in both samples being compared).

1.2.7. Evaluation [Steps 1100–1340]

In previous stages of the handwriting examination process, the FDE determined that the writing to be compared is

- Sufficiently clear and detailed,
- In a character set with which the FDE is comfortable,
- Of sufficient amount and complexity for comparison,
- Actually comparable (i.e., comprised of the same allographs), and
- Internally consistent.

With the combination of observed characteristics in the questioned and known writing samples now classified as either similarities or dissimilarities, the FDE determines the significance of those features. If similarities and no differences are observed, the questioned and known samples may have a common writer, a different writer copying the known writer's handwriting features, or a chance match between different writers. Therefore, in assessing the significance of handwriting characteristics, the FDE must consider (1) how often features as similar as those observed arise in handwriting specimens from the same person (persistence and frequency of features) and (2) how often features as similar as those observed arise in the handwriting from different people (either from chance match or simulation). Section 2.3 expands the discussion of feature interpretation.

Dissimilarities can be expected if different people wrote the questioned and known documents but can also be observed even if the known writer wrote the questioned documents. For this reason, the FDE considers several internal and external factors, as outlined in box 1.1, in

determining whether a feature dissimilarity indicates a different writer or is the product of intra-writer variation.

Box 1.1: Factors to consider in evaluating dissimilarities[106]

- Number and nature of specimens, including whether or not they are contemporaneous
- Whether an individual who might be the writer
  - Has alternative writing styles
  - Is ambidextrous
  - Had a change in physical or mental condition that could influence handwriting features (e.g., health, fractures, fatigue, weakness, nervous, or stress)
  - Was concentrating or not concentrating while writing
  - Was trying to disguise or deliberately change their handwriting
  - Was affected by the use or withdrawal of drugs, alcohol, medication, etc.
- Environmental conditions under which the writings were made (e.g., in a moving vehicle)
- Writing instrument and its quality/working order
- Position of the writer, including stance
- Writing surface

The FDE determines if each compared writing set contains enough habitual, distinctive features characteristic of one writer. These features may be similar or dissimilar between the writing sets. Specifically, the FDE considers whether the writing set contains enough meaningful characteristics to express an opinion about writership. If the answer is no, then the FDE will give an inconclusive opinion regarding writership of the items being compared.

If the answer is yes, and the FDE has not yet considered possible manipulation of the document. Action should be taken at this stage to rule out manipulation, particularly if it is a non-original document. For example, in these cases, manipulation is usually in the form of "cut and paste" entries. Figure 1.7 shows two examples of cut and paste manipulation. In larger amounts of continuous writing, the FDE may determine manipulation if there are repeated superimposable entries of letters, letter combinations, or words between the compared writings sets. The writing under examination will lack normal variation and suggest a manipulated document.

[106] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 51–55.

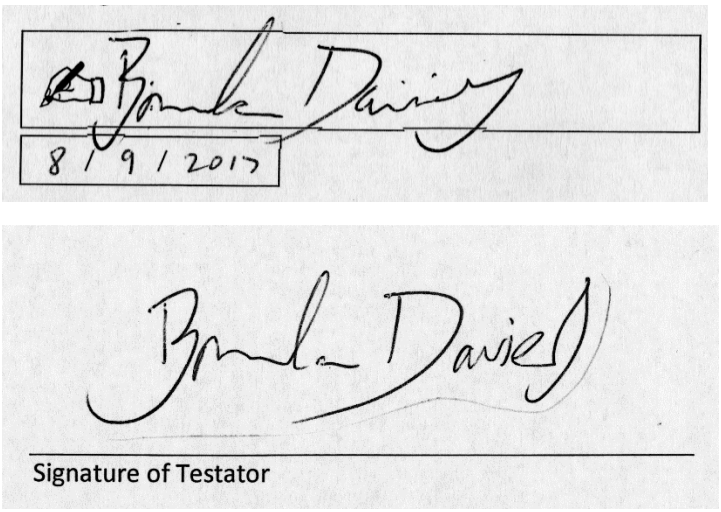


Figure 1.7: Cut and paste manipulation of signatures on non-original documents

The top example shows inconsistencies in the box lines around the signature. The bottom example shows shadowing around the signature caused by cut and paste insertion.

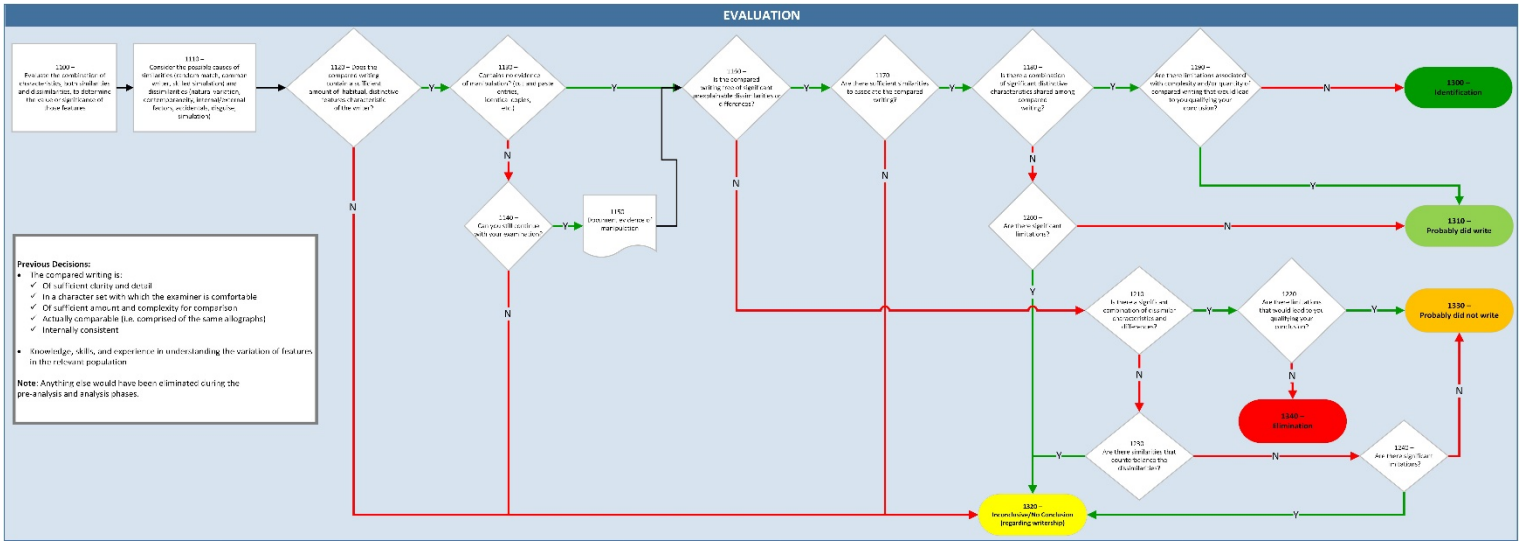
Other forms of manipulation may result in different types of evidence observable in the document, but alterations and manipulations are not the focus of this report. In the case of a manipulated document, it may be possible to express an opinion regarding writership of questioned entries. However, this may be of limited use to the case submitter depending on the question of interest, as it will not be possible to determine how the manipulated entries were incorporated into the document. Therefore, the FDE may decide that it is not possible to continue with the examination and render an inconclusive/no conclusion opinion based on the reasoning outlined in the report.

If the observed evidence of manipulation does not halt the examination process, that evidence is documented, and the examination continues. The process also continues the same way if there is no evidence of manipulation.

Table 1.2 shows the criteria to reach the different levels of identification and exclusion opinions. All other pathways in the process map lead to a report of "no conclusion" regarding writership. By following the process map through the evaluation phase, the relevant decision boxes leading to each conclusion will be completed. The gray shading in table 1.2 indicates that these decision boxes do not appear in the pathway for that conclusion. For certain conclusions, there may be more than one pathway.

Table 1.2: Criteria based on current process map for reaching the different levels of opinion

Are the compared writings free of significant unexplainable dissimilarities or differences?	Are there sufficient similarities in handwriting characteristics to associate the compared writing sets?	Is there a combination of significant, distinctive characteristics shared between the writing sets?	Is there a significant combination of dissimilar characteristics and differences that would point toward different writers?	Are there similarities in handwriting characteristics that counterbalance the dissimilarities?	Are there limitations associated with the complexity or quality of the writing sets that would qualify the conclusion?	Are there significant limitations in the compared material?	OPINION
Yes	Yes	Yes			No		Identification
Yes	Yes	No				No	Probably did write
Yes	Yes	Yes			Yes		
All other pathways within the process map will lead to an "Inconclusive" opinion							Inconclusive
No			No	No		No	Probably did not write
No			Yes		Yes		
No			Yes		No		Elimination



The questions to consider in evaluating the observed handwriting characteristics are described in the following list:

- Are the compared writings free of significant[107] unexplainable dissimilarities or differences? Box 1.1 lists factors to consider when evaluating dissimilarities.
- If so, are there sufficient similarities[108] in handwriting characteristics to associate the compared writing sets?
- If so, is there a combination of significant, distinctive characteristics shared between the writing sets?
- Is there a significant combination of dissimilar characteristics and differences that would point toward different writers?
- If the observed combination of dissimilar or different characteristics is not significant, are there similarities in handwriting characteristics that counterbalance the dissimilarities? In other words, could the observed evidence be caused by the questioned sample having been written by the known writer or by someone else?
- Are there limitations associated with the complexity or quality of the writing sets that would qualify the conclusion?
- Are any limitations significant? These limitations may include non-original documents, low complexity, or a relatively small amount of handwriting for comparison.

Typically, the FDE's task is to ascertain whether known and questioned writings are associated—whether they are written by the same or different individuals. At the end of the evaluation stage, the FDE expresses an opinion indicating subjective confidence in the process outcome. The five opinions given in the process map (identification, probably did write, inconclusive, probably did not write, and elimination) may not map directly onto a given FDE's opinion levels, but they do represent a general opinion scale commonly used in FDE proficiency tests. Sections 1.3 and 3.3 provide further discussion of opinion scales.

At this point, the FDE documents the findings and the basis for the opinion. The FDE determines if all the submitter's questions have been answered. If not, then appropriate additional examinations are conducted, or the FDE documents the reasons why they were not. The FDE then drafts a preliminary report.

1.2.8. Case Review and Report Finalization [Steps 1400–1700]

The written report by the FDE may then be reviewed according to laboratory policy. The types of reviews undertaken are usually technical and administrative, with independent re-examination also possible. Section 4.2.3.2 describes these and other types of reviews. In cases where the FDE and reviewer disagree, the conflict will be resolved according to the laboratory's conflict resolution policy. This disagreement and resolution must be documented in the case notes.

[107] Note that this does not imply statistical significance but a measure of importance.

[108] Sufficient similarities would be those the FDE would not expect to see because of a chance match.

After the report has been reviewed and amended (if necessary), the laboratory notifies the submitter and transmits the report. Private FDEs may provide a verbal report and ask if a written report is needed. If a verbal or written report is not required, the FDE documents the examination results and opinions in the case notes. See section 3.4 for further discussion on reporting requirements.

The examination then concludes. It may be restarted if other documents are submitted or additional examinations are requested.

1.3. FDE Opinions

An FDE's opinion regarding writership can be thought of as expressing a subjective probability[109] for the proposition[110] of a common source. In the conventional approach, this is expressed via a verbal scale.[111] The scales FDEs use to express their opinions currently range from identification (the person who wrote the questioned writings is the same person who wrote the known writings) to elimination (the person who wrote the questioned writing is not the same person who wrote the known writings). These opinions may be reported in terms of ordinal scales ranging from as few as three to as many as thirteen levels.[112] The formation and use of any scale is ultimately left to the laboratory or FDE.

The Scientific Working Group for Forensic Document Examination (SWGDOC) published *Standard Terminology for Expressing Conclusions of Forensic Document Examiners*,[113] summarized in table 1.3, which provides nine opinions (and associated descriptions) that an FDE may express. The FBI laboratory uses five categories that collapse SWGDOC opinions 2 through 4 into "may have (qualified opinion)" and opinions 6 through 8 into "may not have (qualified opinion)."[114] Forensic document examination proficiency test provider Collaborative Testing Services (CTS) uses another five-category scale. All FDEs who undertake these proficiency tests have to use this opinion scale, regardless of what scale they use for reporting their usual casework. An even simpler scale treats the FDE's decision or judgement as binary (yes/no)—a positive association (the questioned writing was produced by the subject) or a negative association (the question writing was not produced by the subject)—but judgement is sometimes reserved by stating that the information in the samples is inconclusive.

[109] The concept of subjective or personal probability is discussed in chapter 2, appendix 2A.

[110] Throughout this report, the terms proposition and propositions are used to denote forensically relevant hypotheses.

[111] Although the Working Group recognizes that the SWGDOC *Standard Terminology* is expressly not to be used as a scale, we are applying the term scale to these conclusion terminology guides based on the concept or definition of an ordinal scale. An ordinal scale is one that has ordered categories. Compare this with a nominal scale, which just has named (mutually exclusive) categories; an interval scale, in which the distance between the categories is known and meaningful; and a ratio scale, which has known distances between the categories and also an absolute zero that is meaningful (hence, a meaningful ratio can be constructed from two values on a ratio scale). These levels of measurement exist within a hierarchy, from low to high: nominal, ordinal, interval, and ratio.

[112] M. L. Merlino et al., *Validity, Reliability, Accuracy, and Bias in Forensic Signature Identification*, National Institute of Justice (2015), https://www.ncjrs.gov/pdffiles1/nij/grants/248565.pdf. A discussion on the range of opinions expressed by document examiners is also presented in S. C. Leung and Y. L.. Cheung, "On Opinion," *Forensic Science International* 42 (1989).

[113] *SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners,* Version 2013-2, (Scientific Working Group for Forensic Document Examination (SWGDOC), 2013).

[114] Harrison, Burkes, and Seiger, "Handwriting Examination: Meeting the Challenges of Science and the Law."

Table 1.3: Summary of SWGDOC Standard Terminology for Expressing Conclusions of FDEs

1. Identification (definite conclusion of identity)	The highest degree of confidence expressed by FDEs in handwriting comparisons. The FDE has no reservations whatsoever, and although prohibited from using the word "fact," the FDE is certain, based on evidence contained in the handwriting, that the writer of the known material actually wrote the questioned document.
2. Strong probability (highly probable, very probable)	The evidence is very persuasive, yet some critical feature or quality is missing so that an identification is not in order; however, the FDE is virtually certain that the questioned and known writings were written by the same individual.
3. Probable	The evidence contained in the handwriting points rather strongly toward the questioned and known writings having been written by the same individual; however, it falls short of the "virtually certain" degree of confidence.
4. Indications (evidence to suggest)	A body of writing has few features of significance for handwriting comparison purposes, but those features are in agreement with another body of writing.
5. No conclusion (totally inconclusive, indeterminable)	This is the zero point of the confidence scale. It is used when there are significantly limiting factors, such as disguise in the questioned or known writing or a lack of comparable writing, and the FDE does not have an opinion one way or another.
6. Indications did not	This carries the same weight as the "indications" term; that is, a body of writing has few features of significance for handwriting comparison purposes, but those features are in disagreement with another body of writing.
7. Probably did not	The evidence points rather strongly against the questioned and known writings having been written by the same individual, but like the probable range above, the evidence is not quite up to the "virtually certain" range.
8. Strong probability did not	This carries the same weight as strong probability on the identification side of the scale; that is, the FDE is virtually certain that the questioned and known writings were not written by the same individual.
9. Elimination	This, like the definite conclusion of identity, is the highest degree of confidence expressed by the document FDE in handwriting comparisons. By using this expression, the FDE denotes no doubt in his or her opinion that the questioned and known writings were not written by the same individual.

Table 1.4 summarizes the particular conclusions within these various opinion scales, which are used in forensic handwriting examination practice, testing, and research. Although some terms in the different scales are similar, how these conclusions are expressed in reports—both between

users of the same scale and between users of different scales—may vary. Box 1.2 provides examples of different expressions of an identification conclusion.

Box 1.2: Examples of identification conclusion wording used by FDEs in reports

In my opinion, the questioned handwriting on item 1 was written by the writer of the known handwriting appearing on items 2 and 3.

John Doe was identified as the writer of the questioned material.

It was determined that John Doe prepared the questioned writing on item 1.

The item 1 questioned writing and the item 2 known writing were prepared by the same individual, identified as John Doe.

Table 1.4: Examples of FDE opinions

A	B	C	Modular Approach	D	E	F
• Identification • Inconclusive • Elimination	• Was written by • Was probably written by (some degree of identification) • Cannot be identified or eliminated • Was probably not written by (some degree of elimination) • Was not written by	• Identification • May have (qualified opinion) • Inconclusive • May not have (qualified opinion) • Elimination	• Evidence provides very strong support for H_1^* over H_2^* • Evidence provides qualified support for H_1 over H_2 • Evidence provides approximately equal support for H_1 and H_2/no conclusion • Evidence provides qualified support for H_2 over H_1 • Evidence provides very strong support for H_2 over H_1	• Identification • Probably did write • Indications did write • Inconclusive/no conclusion • Indications did not write • Probably did not write • Elimination	• Extremely strong support (written by) • Strong support (written by) • Moderate support (written by) • Limited support (written by) • Inconclusive • Limited support (not written by) • Moderate support (not written by) • Strong support (not written by) • Extremely strong support (not written by)	• Identification (definite conclusion of identity) • Strong probability (highly probable, very probable) • Probable • Indications (evidence to suggest) • No conclusion (totally inconclusive, indeterminable) • Indications did not • Probably did not • Strong probability did not • Elimination

Notes:

A: Conclusions that are often required by handwriting studies.

B: Five-point opinions used by CTS.

C: Five-point opinions used by the FBI.

Modular Approach: Modular approach outlined in Found and Bird.[115]

D: Seven-point opinions.

E: Nine-point opinions defined by the European Network of Forensic Handwriting Experts in their Collaborative Exercise program.

F: Nine-point opinions outlined by SWGDOC.

* H_1 and H_2 are used to denote two mutually exclusive hypotheses. For example, H_1 = the same writer wrote the known and questioned writing; H_2 = the questioned writing was written by someone other than the person who wrote the known writing.

[115] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 71.

2. Interpretation and Technology

Introduction and Scope

A forensic handwriting examination involves a series of decisions that depend on careful observation and interpretation of the handwriting evidence. Given the human element of this interpretation process, it also requires awareness and mitigation of the potential for contextual bias. With this in mind, the first section of this chapter focuses on the nature of cognitive bias as it pertains to evidence interpretation and strategies for its mitigation.

The second section of this chapter explores the concepts of error,[116] reliability, and validity. These concepts are particularly important to consider in the study of human factors in handwriting examination because the FDE is the main "instrument" in the examination process. Furthermore, establishing reliability and validity of a technique is pertinent to the court's determination of evidence admissibility.

The third section of this chapter discusses the role of human factors in selecting, weighting, and interpreting features in handwriting evidence, and the statistical approach to evidence interpretation. The final section of this chapter discusses automated systems and technology designed to reduce error in forensic handwriting comparisons. This discussion includes the advantages and limitations of such systems.

2.1. Cognitive Bias

As long as a person is the main instrument of analysis and interpretation in forensic impression and pattern evidence disciplines, the strengths and limitations of human cognition will be central to forensic casework. Although there is nothing inherently wrong with subjective judgements, there may be a higher likelihood of task-irrelevant information affecting the examination. Although quantitative measurements are also human-dependent to some degree and are not immune to the effects of task-irrelevant or other contextual information, the impact may be more transparent. Not all handwriting and other pattern examinations are trivially obvious—if they were, there would be little need for trained experts—and so human cognition plays a critical role in the judgements and performance of FDEs and other examiners. For example, in LPE, not only is there inter-examiner variability in the analysis, interpretation, and conclusion on the same prints, but the same LPE may result in a different conclusion upon re-examination of the same prints.[117] There is no manifest reason not to assume that the same type of variation is likely to hold true among FDEs.

[116] See A. M. Christensen et al., "Error and Its Meaning in Forensic Science," *Journal of Forensic Sciences* 59, no. 1 (Jan 2014), https://doi.org/10.1111/1556-4029.12275.

[117] I. E. Dror et al., "Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison," *Forensic Science International* 208, no. 1-3 (May 20 2011), https://doi.org/10.1016/j.forsciint.2010.10.013, https://www.ncbi.nlm.nih.gov/pubmed/21129867.

A robust body of research examines factors that affect human interpretation, judgement, and decision making.[118] People are predisposed to economize cognitive efforts by using shortcuts like heuristics—mental "rules of thumb" that allow us to solve problems without taxing the brain. These shortcuts lead to cognitive bias, which is neither conscious nor intentional; it is a trade-off that allows humans to quickly and efficiently process large amounts of information in a short time.[119] For example, Tversky and Kahneman[120] discussed various forms of cognitive bias resulting from the "availability heuristic." One such example is bias because of the effectiveness of a search set:

> Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with r or that r is the third letter? People approach this problem by recalling words that begin with r (road) and words that have r in the third position (car) and assess the relative frequency by the ease with which words of the two types come to mind. Because it is much easier to search for words by their first letter than by their third letter, most people judge words that begin with a given consonant to be more numerous than words in which the same consonant appears in the third position. They do so even for consonants, such as r or k, which are more frequent in the third position than in the first.[121]

[118] For example: S. Chaiken, A. Liberman, and A. H. Eagly, "Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context," in *Unintended Thoughts*, ed. J. S. Uleman and J. A. Bargh (New York: The Guilford Press, 1989); Dieter Frey, "The Effect of Negative Feedback About Oneself and Cost of Information on Preferences for Information About the Source of this Feedback," *Journal of Experimental Social Psychology* 17, no. 1 (1981), https://doi.org/10.1016/0022-1031(81)90005-6; Dieter Frey, "Postdecisional Preference for Decision-Relevant Information as a Function of the Competence of its Source and the Degree of Familiarity with this Information," *Journal of Experimental Social Psychology* 17, no. 1 (1981), https://doi.org/10.1016/0022-1031(81)90006-8; Dieter Frey and Dagmar Stahlberg, "Selection of Information after Receiving more or Less Reliable Self-Threatening Information," *Personality and Social Psychology Bulletin* 12, no. 4 (1986), https://doi.org/10.1177/0146167286124006; Dieter Frey, "Recent Research on Selective Exposure to Information," *Advances in Experimental Social Psychology* 19 (1986), https://doi.org/10.1016/S0065-2601(08)60212-9; Dieter Frey and Marita Rosch, "Information Seeking after Decisions: The Roles of Novelty of Information and Decision Reversibility," *Personality and Social Psychology Bulletin* 10, no. 1 (1984), https://doi.org/10.1177/0146167284101010; D. Frey and S. Schulz-Hardt, "Confirmation Bias in Group Information Seeking and Its Implications for Decision Making in Administration, Business and Politics," in *Social Influence in Social Reality: Promoting Individual and Social Change*, ed. F. Butera and G. Mugny (2001); D. Frey, D. Stahlberg, and A. Fries, "Information Seeking of High- and Low-Anxiety Subjects After Receiving Positive and Negative Self-Relevant Feedback," *Journal of Personality* 54, no. 4 (Dec 1986), https://doi.org/10.1111/j.1467-6494.1986.tb00420.x; Dieter Frey and Robert A. Wicklund, "A Clarification of Selective Exposure," *Journal of Experimental Social Psychology* 14, no. 1 (1978), https://doi.org/10.1016/0022-1031(78)90066-5; E. Jonas et al., "Confirmation Bias in Sequential Information Search after Preliminary Decisions: an Expansion of Dissonance Theoretical Research on Selective Exposure to Information," *Journal of Personality and Social Psychology* 80, no. 4 (Apr 2001), https://doi.org/10.1037//0022-3514.80.4.557; Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* 2, no. 2 (1998), https://doi.org/10.1037/1089-2680.2.2.175; M. E. Oswald and S. Grosjean, "Confirmation Bias," in *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*, ed. R. F. Pohl (Hove and NY: Psychology Press, 2004).

[119] J. McClelland and D. Rumelhart, "An Interactive Activation Model of Context Effects in Letter Perception: Part 1, an Account of Basic Findings," *Psychological Review* 88, no. 2 (2011); T. D. Wilson and N. Brekke, "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations," *Psychological Bulletin* 116, no. 1 (Jul 1994), https://doi.org/10.1037/0033-2909.116.1.117.

[120] Amos Tversky and Daniel Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* 5, no. 2 (1973), https://doi.org/10.1016/0010-0285(73)90033-9. See also Daniel Kahneman and Amos Tversky, "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology* 3, no. 3 (1972): 41, https://doi.org/10.1016/0010-0285(72)90016-3.

[121] Tversky and Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," 11.

Scholars have begun to discuss the potential for bias in forensic examinations extensively.[122] Risinger, Saks, Thompson, and Rosenthal argued that "the most obvious danger in forensic science is that an FDE's observations and conclusions will be influenced by extraneous, potentially biasing information."[123] This may result in confirmation bias, which is the tendency to search for or interpret new information in a way that confirms one's preconceptions and avoids information and interpretations that contradict prior beliefs.[124]

Festinger believed that selective attention to information occurs only if the decision is made under free choice and if the person is committed to the decision.[125] He predicted that under specific conditions, people actively seek information that either bolsters their argument or produces easily refutable discordant findings. By doing so, they build a case for their decisions by attending to information that either supports their argument (selective attention) or easily disconfirms alternative explanations (selective information seeking).

Frey and colleagues found that people usually prefer supporting information if they have decided voluntarily on a particular alternative.[126] Confirmation bias is amplified if commitment is heightened,[127] the sources of information are experts rather than lay people,[128] or the decision is irreversible.[129] Confirmation bias has also been found to be stronger in anxious individuals[130] and increases if there are heightened costs associated with the information search (e.g., financial cost/price per additional source).[131]

---

[122] For example, see: Itiel E. Dror, "The Paradox of Human Expertise: Why Experts Can Get It Wrong," in *The Paradoxical Brain*, ed. N. Kapur (Cambridge: Cambridge University Press, 2011); Itiel E. Dror et al., "When Emotions Get the Better of Us: the Effect of Contextual Top-Down Processing on Matching Fingerprints," *Applied Cognitive Psychology* 19, no. 6 (2005), https://doi.org/10.1002/acp.1130; Itiel E. Dror and David Charlton, "Why Experts Make Errors," *Journal of Forensic Identification* 56, no. 4 (2006); I. E. Dror, D. Charlton, and A. E. Peron, "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications," *Forensic Science International* 156, no. 1 (Jan 6 2006), https://doi.org/10.1016/j.forsciint.2005.10.017, https://www.ncbi.nlm.nih.gov/pubmed/16325362; Itiel E. Dror and J. L. Mnookin, "The Use of Technology in Human Expert Domains: Challenges and Risks Arising from the Use of Automated Fingerprint Identification Systems in Forensic Science," *Law, Probability and Risk* 9, no. 1 (2010), https://doi.org/10.1093/lpr/mgp031; Itiel E. Dror and S. A. Cole, "The Vision in 'Blind' Justice: Expert Perception, Judgment, and Visual Cognition in Forensic Pattern Recognition," *Psychonomic Bulletin & Review* 17, no. 2 (Apr 2010), https://doi.org/10.3758/PBR.17.2.161, https://www.ncbi.nlm.nih.gov/pubmed/20382914; I. E. Dror et al., "The Impact of Human-Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts," *Journal of Forensic Sciences* 57, no. 2 (Mar 2012), https://doi.org/10.1111/j.1556-4029.2011.02013.x; William C. Thompson, "What Role Should Investigative Facts Play in the Evaluation of Scientific Evidence?," *Australian Journal of Forensic Sciences* 43, no. 2-3 (2011), https://doi.org/10.1080/00450618.2010.541499; Itiel E. Dror and R. Rosenthal, "Meta-Analytically Quantifying the Reliability and Biasability of Forensic Experts," *Journal of Forensic Sciences* 53, no. 4 (Jul 2008), https://doi.org/10.1111/j.1556-4029.2008.00762.x.

[123] D. M. Risinger et al., "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion," *California Law Review* 90, no. 1 (2002): 9, https://heinonline.org/HOL/LandingPage?handle=hein.journals/calr90&div=10&id=&page=.

[124] Oswald and Grosjean, "Confirmation Bias."; Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises."

[125] L. Festinger, *A Theory of Cognitive Dissonance* (Stanford, CA: Stanford University Press, 1957).

[126] Frey, "Recent Research on Selective Exposure to Information."; Frey and Wicklund, "A Clarification of Selective Exposure."; Frey and Schulz-Hardt, "Confirmation Bias in Group Information Seeking and Its Implications for Decision Making in Administration, Business and Politics."

[127] Frey, Stahlberg, and Fries, "Information Seeking of High- and Low-Anxiety Subjects After Receiving Positive and Negative Self-Relevant Feedback."

[128] Frey, "The Effect of Negative Feedback About Oneself and Cost of Information on Preferences for Information About the Source of this Feedback."

[129] Frey, "Postdecisional Preference for Decision-Relevant Information as a Function of the Competence of its Source and the Degree of Familiarity with this Information.".

[130] Frey, Stahlberg, and Fries, "Information Seeking of High- and Low-Anxiety Subjects After Receiving Positive and Negative Self-Relevant Feedback."

[131] Frey, "The Effect of Negative Feedback About Oneself and Cost of Information on Preferences for Information About the Source of this Feedback."

Several factors, including time pressure or high complexity,[132] appear to exacerbate a confirmation bias before making a final decision. For example, Frey et al.[133] found that such circumstances may override the person's desire (or ability) to critically test the primary conclusion against all available alternatives. Confronted with evidence backlogs, time pressures, or other difficult conditions, decision makers may subconsciously engage in cognitive behaviors that allow for diminished cognitive effort (e.g., selective attention or selective information seeking).

Another factor that can exacerbate confirmation bias is the strength of the person's own opinions or beliefs. Edwards and Smith[134] reported that supporting information is perceived to be more credible and valid (better) than information that refutes what one knows. Differentially evaluating supporting and conflicting arguments seems to elicit a preference for supporting information, even without motivation to have one's preferences or prior decisions confirmed.

Finally, the need to justify a decision to others (e.g., supervisors, colleagues) can result in an "impression motivation."[135] Here, people may seek out disproportionately supporting information because this information helps justify a decision.[136]

Although there is currently limited research about this issue's impact on handwriting examination specifically,[137] bias has been identified as an issue in many other forensic disciplines.[138] Therefore, the Working Group does not assume FDEs are immune from cognitive and contextual bias.

In recognizing that bias is a legitimate cause for concern in forensic science, several large reports have called for forensic laboratories to mitigate its potential negative effects. A committee of the

---

[132] Jonas et al., "Confirmation Bias in Sequential Information Search after Preliminary Decisions: an Expansion of Dissonance Theoretical Research on Selective Exposure to Information."; Frey and Rosch, "Information Seeking after Decisions: The Roles of Novelty of Information and Decision Reversibility."

[133] D. Frey et al., Information seeking under suboptimal conditions: The importance of time pressure and complexity for selective exposure to information, 2000, University of Munich.

[134] K. Edwards and E. E. Smith, "A Disconfirmation Bias in the Evaluation of Arguments," *Journal of Personality and Social Psychology* 71, no. 1 (1996).

[135] Chaiken, Liberman, and Eagly, "Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context."

[136] Jonas et al., "Confirmation Bias in Sequential Information Search after Preliminary Decisions: an Expansion of Dissonance Theoretical Research on Selective Exposure to Information."

[137] Early work on this issue used trainee examiners; therefore, the generalizability to expert FDEs is unclear. See; Larry S. Miller, "Bias Among Forensic Document Examiners: A Need for Procedural Change," *Journal of Police Science & Administration* 12, no. 4 (1984), https://psycnet.apa.org/record/1985-17219-001. In another study, lay people judged handwriting samples in presence or absence of a confession, see: J. Kukucka and S. Kassin, "Do Confessions Taint Perceptions of Handwriting Evidence? An Empirical Test of the Forensic Confirmation Bias," *Law and Human Behavior* 38, no. 3 (2014).

[138] For examples, see: Dror and Charlton, "Why Experts Make Errors."; Dror, Charlton, and Peron, "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications."; Dror et al., "Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison."; P. A. Fraser-Mackenzie, I. E. Dror, and K. Wertheim, "Cognitive and Contextual Influences in Determination of Latent Fingerprint Suitability for Identification Judgments," *Science & Justice: Journal of the Forensic Science Society* 53, no. 2 (Jun 2013), https://doi.org/10.1016/j.scijus.2012.12.002; J. Kerstholt et al., "Does Suggestive Information Cause a Confirmation Bias in Bullet Comparisons?," *Forensic Science International* 198, no. 1-3 (May 20 2010), https://doi.org/10.1016/j.forsciint.2010.02.007; G. Langenburg, C. Champod, and P. Wertheim, "Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons," *Journal of Forensic Sciences* 54, no. 3 (May 2009), https://doi.org/10.1111/j.1556-4029.2009.01025.x; S. Nakhaeizadeh, I. E. Dror, and R. M. Morgan, "Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias," *Science & Justice: Journal of the Forensic Science Society* 54, no. 3 (May 2014), https://doi.org/10.1016/j.scijus.2013.11.003; N. K. Osborne et al., "Does Contextual Information Bias Bitemark Comparisons?," *Science & Justice: Journal of the Forensic Science Society* 54, no. 4 (Jul 2014), https://doi.org/10.1016/j.scijus.2013.12.005; N. K. Osborne et al., "Bloodstain Pattern Classification: Accuracy, Effect of Contextual Information and the Role of Analyst Characteristics," *Sci Justice* 56, no. 2 (Mar 2016), https://doi.org/10.1016/j.scijus.2015.12.005.

National Research Council (NRC) recommended "standard operating procedures [and] model protocols to minimize, to the greatest extent possible, potential bias . . . in forensic science."[139] The NIST Expert Working Group on latent print analysis noted "the desirability of procedures to help avoid bias."[140] Furthermore, the National Commission on Forensic Science (NCFS) expressed its view that "[f]orensic laboratories should take appropriate steps to avoid exposing analysts to task-irrelevant information through the use of context management procedures detailed in written policies and protocols."[141]

2.1.1. Contextual Bias in Forensic Handwriting Examinations

The remainder of this section focuses on sources of contextual information that could bias an FDE and discusses ways to mitigate the potential effects of bias in casework. Box 2.1 serves as a glossary of terms that relate to bias and contextual information in forensic casework.

Box 2.1: Glossary of terms relating to bias and its management[142]

Bias: A systematic pattern of deviation.

Blind Case: A case developed with the intention of testing the examiner or the examination process and in which the ground truth is known. Critically, the examiner is not aware the case is not genuine.

Blind Declared Case: Blind cases that the examiner knows will be inserted into routine casework. The examiner will not know which cases are blind. See section 4.2.6.4.

Blinding: Systematically shielding an examiner from task-irrelevant contextual information.

Cognitive Bias: A systematic pattern of deviation in human judgement.

Context: The set of circumstances or facts that surround a case.

Context-Manager Model: A type of CIM procedure whereby a forensic expert or administrator filters discipline- and task-irrelevant contextual information from the examiner who is to perform the examination.

Contextual Bias: A type of cognitive bias to denote human judgement being influenced by irrelevant contextual information.

Contextual Information: Knowledge, whether relevant or irrelevant, concerning a particular fact or circumstance related to a case or examination. Contextual information is conceptualized

[139] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, The National Academies Press (Washington, DC, 2009), 24..

[140] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 41.

[141] National Commission on Forensic Science (NCFS), *Views of the Commission: Ensuring that Forensic Analysis Is Based Upon Task-Relevant Information*, Department of Justice (2015), 1, https://www.justice.gov/archives/ncfs/file/818196/download.

[142] Unless otherwise stated, these terms are defined by the Working Group based on the relevant literature and how the terms are used within the context of this report.

in different levels (see sections 2.1.2 to 2.1.6). These levels are ordered with respect to how far removed the information is from the questioned material and the examination.

Contextual Information Management (CIM): Actions to optimize the flow of information to and from a forensic expert to minimize the potential for contextual bias.

Forensic Discipline: A specialized branch or field of forensic science (e.g., handwriting examination, DNA analysis, LPE, bloodstain pattern analysis).

Irrelevant Information: Information that is not pertinent or applicable to the subject, material, or question being considered. The consideration may be broad (i.e., case or discipline level) or specific (i.e., task level).

Linear Sequential Unmasking (LSU): A type of CIM procedure that specifies the optimal order in which forensic experts should examine the unknown material (e.g., questioned writing) and reference material (e.g., known writing) to conduct a comparison. The experts must examine and document the unknown material before being exposed to the reference material, therefore working from the evidence to the suspect.[143] The term LSU has been coined by Dror and colleagues[144] to stress that the examiner is not allowed unlimited back and forth access between the questioned and known material. LSU follows the same basic principles of sequential unmasking; however, it also requires examiners to specify a level of confidence in their opinion regarding the material under examination.[145]

Relevant Information: Information that is pertinent and applicable to the subject, material, or question being considered. The consideration may be broad (i.e., case or discipline level) or specific (i.e., task level).

Task: A piece of work to be undertaken.

The growing appreciation of the conditions under which cognitive bias can arise in forensic science has spurred the development and implementation of practical solutions to strengthen the reliability and admissibility of the forensic evidence. Contextual information management (CIM) aims to minimize exposure to task-irrelevant information while still allowing the FDE to access information that is relevant to their task.[146] The Working Group recommends the adoption of

[143] Krane et al., "Sequential Unmasking: a Means of Minimizing Observer Effects in Forensic DNA Interpretation."

[144] Dror et al., "Letter to the Editor—Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making," 1112. "Sequential unmasking allows unlimited and unrestricted changes to the evidence once exposed to the reference material. We believe it is important to impose limits and restrictions for when examiners are permitted to revisit and alter their initial analysis of trace evidence. The analysis of traces is most objective when the examination is 'context free'—that is, prior to exposure to the known reference samples. However, seeing the reference samples could alert the examiner to a possible oversight, error, or misjudgment in the analysis of the trace evidence. Here, we seek to strike a balance between restrictive procedures that forbid analysts from changing their opinion and those that allow unlimited and unrestricted changes. The requirement that changes be documented does not eliminate the possibility that such changes arose from bias—it only makes that possibility more transparent."

[145] Because the features that must be considered in a handwriting case are generally not defined before the case, taking a strict approach to LSU in handwriting examination could result in a loss of evidential strength. This is further discussed in section 2.1.3.

[146] R. D. Stoel et al., "Minimizing Contextual Bias in Forensic Casework," in *Forensic Science and the Administration of Justice: Critical Issues and Directions*, ed. Kevin Strom and Matthew Hickman (Thousand Oaks: SAGE Publications, Inc, 2014), 67–86; E. J. Mattijssen et al., "Implementing Context Information Management in Forensic Casework: Minimizing Contextual Bias in Firearms Examination," *Science & Justice: Journal of the Forensic Science Society* 56, no. 2 (Mar 2016), https://doi.org/10.1016/j.scijus.2015.11.004, https://www.ncbi.nlm.nih.gov/pubmed/26976470.

CIM for handwriting examination to minimize FDE exposure to task-irrelevant, potentially biasing contextual information at various stages of forensic work. The idea of managing contextual information in forensic handwriting examination casework is not new.[147] Examples of CIM will be discussed in the following sections.

Understanding how different sources of contextual information affect forensic casework can help mitigate the potential negative effects of bias arising from exposure to this information.[148] Figure 2.1, adapted from Dror,[149] presents a graphical representation of seven levels (i.e., sources) of contextual information. As each level increases in number, it represents greater departure from the material in question (e.g., questioned handwriting). Level 1 (described in section 2.1.2) contains information obtained from the questioned material itself, and Levels 2 through 7 (described in sections 2.1.3 through 2.1.6) subsequently contain information that is more remote from the questioned material.

2.1.2. Level 1 Contextual Information

Level 1 contextual information pertains to the questioned material. It is all the information contained in the questioned material separate from the handwriting features (e.g., type of ink and paper, and meaning of the words). Although this information might be task-relevant at some point in the examination, it is generally task-irrelevant when assessing the handwriting features (see section 3.4.1).

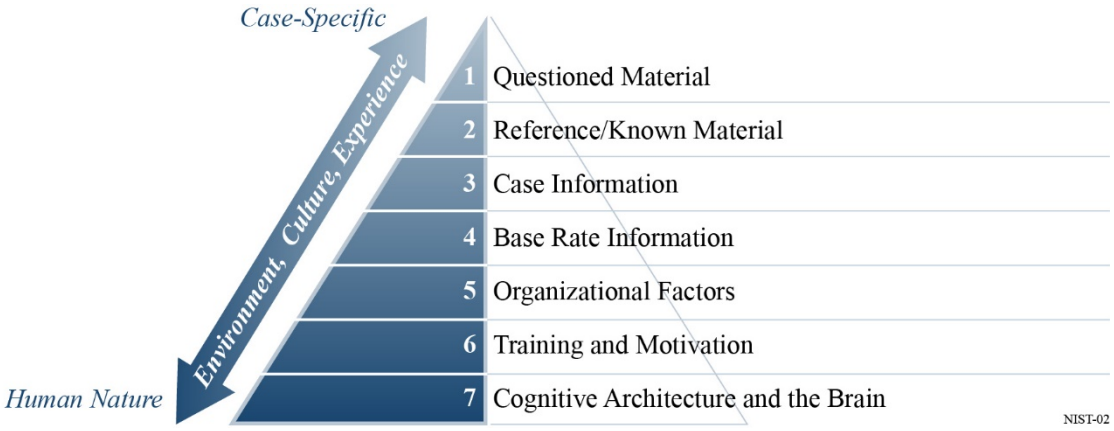


Figure 2.1: Taxonomy of seven sources of contextual information in forensic examinations[150]

Level 1 contextual information is generally difficult to manage because it is inherent in the evidential material and often cannot be easily separated from the handwriting itself. One potentially biasing aspect of Level 1 contextual information is the content and meaning of the

[147] B. Found and J. Ganas, "The Management of Domain Irrelevant Context Information in Forensic Handwriting Examination Casework," *Science & Justice: Journal of the Forensic Science Society* 53, no. 2 (Jun 2013), https://doi.org/10.1016/j.scijus.2012.10.004, https://www.ncbi.nlm.nih.gov/pubmed/23601722.

[148] Itiel E. Dror, "Human Expert Performance in Forensic Decision Making: Seven Different Sources of Bias," *Australian Journal of Forensic Sciences* 49, no. 5 (2017), https://doi.org/10.1080/00450618.2017.1281348.

[149] Dror, "Human Expert Performance in Forensic Decision Making: Seven Different Sources of Bias."

[150] Figure adapted from Dror, "Human Expert Performance in Forensic Decision Making: Seven Different Sources of Bias."

written words. In principle, parts of the evidential material that convey meaning could be removed or presented in a manner that obscures the meaning. However, any CIM of Level 1 contextual information requires careful consideration to balance the need to disguise or remove the potential source of bias and the loss of evidentiary information. Many FDEs, for instance, do not favor using digital scans of questioned documents or the practice of using only part of the available handwriting. Whether that is a legitimate concern should be the topic of future studies.

2.1.3. Level 2 Contextual Information

Level 2 contextual information pertains to the reference material (i.e., known documents). Similar to Level 1 contextual information, the meaning of the words in course-of-business documents, collected as known samples, may subconsciously bias the examiner. In addition, because handwriting examination requires a comparison between the questioned and known handwriting, the features contained in one could influence the selection and interpretation of the features contained in the other.

If FDEs start with the known material, their subsequent analysis of the questioned material could be biased by the information contained in features of known material. That is, features in the questioned material similar to features in the known material could be given more weight than they otherwise would have, and dissimilar features could be ignored or given less weight. By proceeding in this way, FDEs are working from the suspect to the evidence—a potentially dangerous method that should be avoided.

Therefore, as a practical matter, FDEs should always analyze the questioned material to determine which features are present and absent before moving to their examination of the known material (steps 100–230 in the process map). This sentiment can be found in early writings on the subject in 1954 when Böttcher[151] stressed the importance of such an approach in forensic handwriting examination. Dror et al. present a detailed "linear sequential unmasking (LSU)" approach for minimizing bias because of contextual information.[152] However, there has been little discussion of LSU in the context of forensic handwriting examination.

In contrast, LSU is an integral part of LPE. It lies at the core of the ACE-V[153] methodology (analysis, comparison, evaluation, and verification) of friction ridge prints. In this workflow, the latent print examiner must annotate the features of the questioned print expected to be useful in the later comparison before seeing the prints from a known suspect. Other forensic laboratories, such as the Netherlands Forensic Institute and the Dutch National Police, also employ LSU as a standard working procedure for fingerprint and DNA evidence.[154] Once again, the FDE begins with the evidence at hand before being exposed to or working with the reference material.

LSU is appropriate for handwriting examination, but unlike the predefined features in LPE or DNA analysis, the features that must be taken into account in a handwriting case are generally

---

[151] C. .J. F. Böttcher, "Theory and Practice of Forensic Handwriting Comparison," *Tijdschift voor Strafrecht* 63 (1954).

[152] Dror et al., "Letter to the Editor—Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making."

[153] M. Triplett and L. Cooney, "Etiology of ACE-V and Its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing," *Journal of Forensic Identification* 56, no. 3 (2006), https://pdfs.semanticscholar.org/0b71/cba100dce3da6266d22df6045c37faf4ee77.pdf.

[154] Stoel et al., "Minimizing Contextual Bias in Forensic Casework."

not defined beforehand. Taking a strict approach to LSU in handwriting examination could result in a loss of evidential strength if not all discriminatory features are identified in the initial examination of the questioned writing and are therefore not considered in the comparison.

Studies are needed to understand the trade-off between discriminatory power, efficiency, and risk of bias in applying LSU to handwriting examinations. Nevertheless, unbiased feature selection is important (see also section 2.3.1), and the management of Level 1 and Level 2 contextual information should not be dismissed based on an efficiency argument.

2.1.4. Level 3 Contextual Information

Level 3 contextual information pertains to all information (oral, written, and behavioral) in a case but is not directly part of the questioned or known material. An FDE might be exposed to Level 3 information via communication with colleagues, the police, or the prosecutor; through written reports, oral discussions, and exchanges; or through nonverbal communication. Some of the available information is important for the forensic expert undertaking the comparison to know (i.e., task-relevant), some may be important for an expert from another discipline (i.e., task-irrelevant for the FDE but task-relevant for examiners in other disciplines), and some is important for the judge or jury but is not relevant to the FDE or examiners in other disciplines (i.e., case-relevant but task- and discipline-irrelevant for the FDE).

The main approach suggested to reduce bias from Level 3 contextual information is to avoid exposure to the information in the first place. As explained by Found and Ganas,[155] an FDE (or another person trained in recognizing task-relevant and task-irrelevant information) can screen the case material so that the FDE who performs the comparison is shielded from the task-irrelevant information. Found and Ganas[156] describe the context-manager model, whereby a context manager removes task-irrelevant information from the case file, leaving FDEs with only the information relevant for the handwriting examination and comparison.

2.1.5. Level 4 Contextual Information

Level 4 contextual information pertains to organization- and discipline-specific "base rate" information that can create an expectation about the outcome of a case. Case work submitted for examination, whether in a criminal or civil case, often undergoes a selection process, and the FDE may be aware of that. For instance, it has been claimed that most evidence presented for forensic evaluation in criminal cases results in a conclusion that associates a suspect with the case.[157] By being aware of such information, FDEs may have a heightened expectation that the evidence is inculpatory, even before the examination has started. Although the base rate has no effect on the actual strength of the evidence, it can bias the FDE toward over- or underestimating the strength of the evidence.

Base rate information may result in a continuing expectation that the evidence under consideration is inculpatory, but the FDE's opinion should be based on the evidence without preconceptions. A mitigating procedure would be to inform FDEs that their case flow will

[155] Found and Ganas, "The Management of Domain Irrelevant Context Information in Forensic Handwriting Examination Casework."

[156] Found and Ganas, "The Management of Domain Irrelevant Context Information in Forensic Handwriting Examination Casework."

[157] Risinger et al., "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion."

include simulated cases with "innocent" writers. As a practical matter, however, creating enough blind cases that the FDEs would perceive as real could be difficult, and expending a great deal of FDE time and effort to blind cases would be costly. However, Stoel et al. note that the psychological effect of knowing that such cases are part of the case flow could be greater than their numerical proportion would suggest.[158] The feasibility and efficacy of inserting declared blind cases into routine cases, therefore, merits study.

2.1.6. Levels 5 to 7 Contextual Information

Level 5 includes a variety of human factors that stem from the organization of the laboratory and its culture (discussed further in chapter 6). Level 6 consists of the training and motivation of the FDEs (discussed further in chapter 5). Level 7 constitutes cognitive architecture and the brain and is intrinsically connected to all human factor issues.[159]

2.1.7. CIM and Task Relevance

According to Risinger,[160] many forensic practitioners claim that their extensive training programs will provide a protective factor against bias; however, he posits that experts "are no more successful in guarding against such distortions by willing them away than any other group ever studied." Training for forensic practitioners should certainly include the topic of cognitive bias but as in other fields of science and medicine,[161] methods that shield FDEs from biasing information will likely be more effective than training alone.

Regardless of which CIM method an analyst employs, the critical determination is the relevance and irrelevance of information to the analyst's task. This may indeed pose challenges for an FDE because handwriting is only one sub-discipline of Questioned Documents. For example, information about ink dating, paper composition, and location of indented writing may not be necessary to the handwriting comparison but may be relevant to other aspects of a case. In most cases, however, items of contextual information can be triaged according to what, when, and to whom it is relevant. Figure 2.2 demonstrates how information might be relevant for a whole case, might only be relevant for one forensic discipline, or more specifically, might only be relevant for one task within that discipline.

At the broadest level, all information relevant to an overall case or investigation falls under the umbrella of case information (figure 2.2, red circle). For example, eyewitness reports, confessions, fingerprint evidence, and handwriting samples are all sources of case information (depending on the case). Who considers that information, and when, are critical elements for reducing bias-related error. For example, a confession is relevant for the overall case (and must be considered by investigators and those deciding on the ultimate issue [e.g., judge, jury]) but should never be considered by forensic scientists drawing opinions from scientific evidence.

---

[158] Stoel et al., "Minimizing Contextual Bias in Forensic Casework."

[159] Dror, "Human Expert Performance in Forensic Decision Making: Seven Different Sources of Bias."

[160] D. M. Risinger, "The NAS Report on Forensic Science: a Glass Nine-Tenths Full (This is About the Other Tenth)," *Jurimetrics* 50, no. 1 (2009): 24.

[161] C. T. Robertson and A. S. Kesselheim, eds., *Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law* (Atlanta, GA: Elsevier, 2016).

Discipline-relevant information (figure 2.2, yellow circle), which lies within the umbrella of case information, might be relevant for one discipline but not another. Those with knowledge of how the case information is relevant to each discipline should manage this information so that an FDE only receives information that falls within their discipline of expertise. For example, an opinion regarding a fingerprint examination (a discipline relevant for latent print analysis) is not relevant to and should never be considered by the expert who conducts the handwriting (or any other) examination.

The relevance of discipline-specific information will further depend on the given task in which the expert is engaging (figure 2.2, green circle). Tasks are the components or pieces of work an examiner undertakes within any given discipline. FDEs are required to engage in numerous tasks within the overall discipline of forensic document examination, and information that might be relevant for one task will not be relevant for another. For example, when conducting an analysis of the questioned writing, knowledge of the features in the known writing is task-irrelevant, even though it is discipline-relevant. When making a comparison between the known and questioned writing, however, knowledge of the features in the known writing becomes task-relevant information.

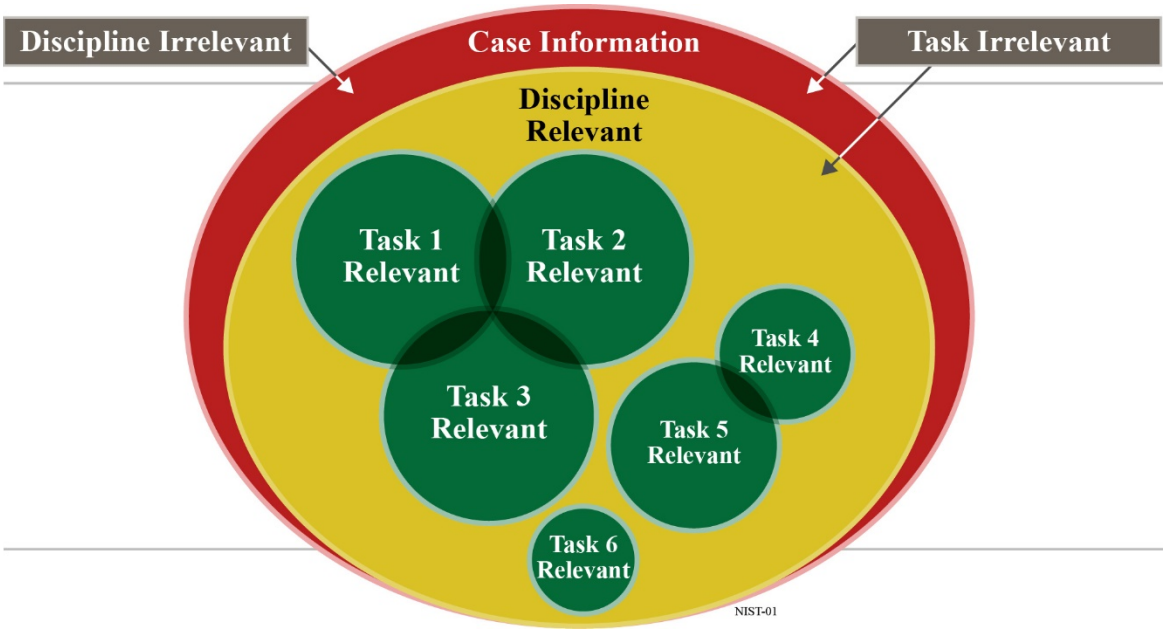


Figure 2.2: Information (ir)relevance as a function of case, discipline, and task

Figure 2.2 highlights that case information can be both discipline-irrelevant and task-irrelevant. Furthermore, some discipline-relevant information can be both task-relevant and task-irrelevant, depending on the task. In practice, a single case may require experts from multiple disciplines (i.e., multiple yellow circles within the red circle), and multiple tasks within the discipline(s) (i.e., multiple green circles within the yellow circles).

Consider a case in which the main question for an FDE is whether a suicide note was written by the deceased or by his non-identical twin brother. According to a police report, the twin brother, who lived in the same household, is in serious financial trouble. Their father, who died of natural causes a week earlier, left an unexpectedly large inheritance to be divided evenly between the twins. The full inheritance would be sufficient to rid the surviving twin brother of his debts. Widely known for his short temper, this twin has two convictions for violent crimes. DNA and a fingerprint matching the living twin brother were found on the suicide note. All this information is in the police report that accompanies a request to the laboratory to examine the suicide note. Along with the suicide note, the police supply some collected handwriting from both brothers and a set of requested samples from the suspected twin. The deceased's handwriting samples consist of several recent shopping lists and a diary.

The information in this case report (i.e., case information) could be critical for the investigator and the trier of fact. All of it (except for the information that the reference material is recent), however, is irrelevant to the comparison of the handwriting and might influence the FDE to arrive at a particular conclusion. Therefore, the FDE who compares the handwriting of the note with the reference material from both twins should not be aware of the suspicion, the financial troubles, the inheritance, the violent behavior, or the DNA and fingerprint evidence (i.e., all discipline- and task-irrelevant information). The task-relevant information is limited to the following: (1) the suicide note, (2) the reference material from both twins, (3) the fact that the reference material and the suicide note are fairly contemporaneous, and (4) the request that the FDE address the proposition of whether the note was written by (a) the deceased, (b) the twin brother, or (c) someone other than the deceased or twin brother.

In some instances, task-relevant information could be biasing. For example, knowing that a person contracted a disease that affects motor skills between the dates that the questioned and known documents were written is certainly relevant. This information could alert the FDE of the possibility that the known writing may not truly represent the writing style that the known writer had contemporaneous with the questioned writing occurring. This information, however, could result in bias if the FDE subconsciously considers the medical information in forming a judgement.

Table 2.1 presents a general framework for deciding when and what type of action should be taken to manage contextual information, according to whether or not information is biasing and relevant.[162] Although in theory, no action is needed for information that is not biasing, it is not always clear when information is biasing. In practice, even though it may be more efficient not to do anything with (i.e., leave in) irrelevant non-biasing information, it may be best to exclude all task-irrelevant information whenever practical.

Table 2.1: Overview of general actions to manage contextual information

	Task-Relevant Information	Task-Irrelevant Information
Biasing	Keep, but take measures.	Shield FDE from this information.
Not Biasing	Use.	Not strictly necessary to shield FDE but shield if possible and efficient.

[162] Stoel et al., "Minimizing Contextual Bias in Forensic Casework."

Other considerations for sole practitioners or small laboratories

Ideally another FDE, or at least a person with similar expertise, should act as the person responsible for the flow of information in a case. This person decides whether CIM is necessary, and if so, what and when information is task-relevant. The actions taken may vary depending on the propositions to be addressed (see section 2.3.2.1), and on the types of contextual information (sections 2.1.2 through 2.1.6) under consideration. The multi-person nature of CIM can pose challenges for sole practitioners or very small teams. Solutions to overcome this challenge include the following:

- Sole practitioners could collaborate with other sole practitioners or laboratories to provide CIM for each other.
- For those working in a multidiscipline laboratory, FDEs could enlist examiners from other disciplines to assist with CIM.
- Administrative staff (where available) could be trained to assist with CIM.
- FDEs could establish clear and transparent agreements with the client regarding what information to give at which moment, before the client submits the case.

In an example taken from firearms examination, Mattijssen et al.[163] described two approaches to shield an examiner from task-irrelevant (primarily Level 3) contextual information. Each approach requires a different list of criteria to determine which information to keep or remove. Approach 1 requires a list of what is classified as task-irrelevant information, which is going to be difficult to identify exhaustively. That is, examiners are shielded only from information that has been identified as task-irrelevant. Approach 2 requires a list of what is classified as task-relevant information, which is much easier to define. Here, firearms examiners are shielded from all verbal and written case information, except for information deemed to be task-relevant.

Mattijssen et al.[164] suggested that the first approach, although intuitively appealing, does not give satisfactory results in practice. Obtaining a complete list of the criteria for task-irrelevant information and implementing these criteria such that every firearms examiner applies them in the same way may be difficult and results in great variation among examiners. The second approach gives more consistent results and is faster.

Over the course of an examination and in preparing the final report, the expert should have gained access to all the task-relevant information. The order in which the FDE receives that information, however, depends on the order in which the tasks were completed. To minimize bias, the tasks must be performed in an order that reduces the potential for cognitive contamination of information between the tasks. Understanding the difference between task and discipline relevance (and irrelevance), and the optimal order of task completion is the cornerstone of LSU.[165]

When developing CIM procedures, laboratories and experts must consider that some experts will perform examinations across multiple disciplines, and many will perform multiple tasks

[163] Mattijssen et al., "Implementing Context Information Management in Forensic Casework: Minimizing Contextual Bias in Firearms Examination."

[164] Mattijssen et al., "Implementing Context Information Management in Forensic Casework: Minimizing Contextual Bias in Firearms Examination."

[165] Dror et al., "Letter to the Editor—Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making."; Krane et al., "Sequential Unmasking: a Means of Minimizing Observer Effects in Forensic DNA Interpretation."

simultaneously within one discipline. Once an FDE has knowledge of information in one discipline or task, it is difficult (if not impossible) for that FDE to simply ignore the information if it is task-irrelevant for subsequent tasks. Here, blind technical reviews or independent re-examinations are particularly important, whereby the reviewer does not know the case information or the original FDE's opinion (see sections 4.2.3.2.2 and 4.2.3.2.3).

In the unsuccessful application of CIM—for example, the FDE was exposed to task-irrelevant information—action may be warranted to determine if the results were adversely affected by the knowledge of this information. The action taken will depend on the specific situation. One option is to redo the CIM and give the complete case to a second or third FDE. All actions (and inactions) should be reported in the case files and/or reports.

For laboratories that routinely perform re-examinations (see section 4.2.3.2), contextual information withheld from the first FDE should also be withheld from the reviewer. The task-irrelevant information includes the conclusion of the first FDE . The re-examination is performed blind to the original conclusion and any information other than what is relevant for review purposes.

Although there is a plethora of experimental research on contextual bias in other forensic disciplines, relatively few studies address forensic handwriting examination. Studies of potential bias and its effects on handwriting examination should consider the following.

- ***Whether some sources of contextual information are more biasing than others.*** Studies should examine the relative contribution of various sources of contextual information (from each of the seven levels) to FDE's opinions.
- ***The optimal order for FDEs to perform their tasks and receive task-relevant information.*** Because contextual information can have a carry-over effect if relevant for one task but irrelevant for another, studies should determine the optimal order for FDEs to (1) perform their tasks and (2) receive contextual information to assist with these tasks.
- ***The efficacy of CIM protocols.*** These studies should address whether or not redacting potentially biasing information during examinations is an effective way of increasing FDE objectivity and reducing bias and which CIM methods are the most effective. These studies could also investigate possible risky shifts (movement toward a more extreme position) or ultra-conservatism in jointly resolved cases.
- ***A cost/benefit analysis of the threshold at which information loss has a greater detrimental impact than risk of bias.*** These studies should address the potential negative impact of shielding FDEs from possible diagnostic information.

Recommendation 2.1: The research community, in collaboration with forensic document examiners, should conduct research to study

- **The impact of various sources of contextual information on forensic handwriting examinations, and**
- **How to balance the risks of bias and information loss with respect to all levels of contextual information.**

Recommendation 2.2: Forensic document examiner laboratories performing handwriting examinations must use a contextual information management protocol, which must be documented within their quality management system.

There is sufficient justification in existing literature to support the immediate implementation of CIM protocols; therefore, the Working Group stresses waiting for the results of Recommendation 2.1 is not necessary to implement Recommendation 2.2. The outcomes from studies that result from Recommendation 2.1. should be used to improve the impact and efficiency of any CIM protocol used.

2.2. Validity and Reliability of Forensic Handwriting Comparisons

This section discusses the scientific basis of validity and reliability pertaining to forensic evidence. The Working Group considered the underlying scientific principles, potential sources of error, the validity and reliability of the analytical methods, and judgements derived from the observational and decisional processes of FDEs. The focus of this section is conceptual, rather than an analysis of the status of validation research.

Both the *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[166] and the Federal Rule of Evidence (FRE) 702 hold that expert testimony be based on methods derived from scientifically valid reasoning and that these methods are applied appropriately to the evidence of a case. However, it is apparent that the forensic community does not apply these putative standards in a uniform manner. Judges, litigants, legal scholars, and forensic scientists may differ in what each views as acceptable scientific validity.[167] The question is whether FDEs can demonstrate the basis for their testimony.

2.2.1. The Appropriateness of the Underlying Principles

The following principles formed the basis for development, application, and interpretation of feature comparison methods in handwriting examination as well as the development of automated handwriting comparison technologies (see section 2.5). First is the principle of individuality: that "no two writers share the same combination of handwriting characteristics given sufficient quantity and quality of writing to compare."[168] The second is the principle "that no two writings by the same person are identical."[169]

[166] *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).

[167] We have located nine district court cases that directly addressed the issue of whether the expert testimony of an FDE is admissible under Daubert and Kumho. No consensus has emerged. Only two courts have found the testimony to be reliable and fully admissible. *United States v. Gricco, No. 01-90*, 2002 WL 746037, 2002 U.S. Dist. LEXIS 7564 (E.D.Pa. 2002); *United States v. Richmond, No. 00-321,*, 2001 WL 1117235, 2001 U.S. Dist. LEXIS 15769 (E.D.La. 2001). Four courts have determined that the FDE's testimony was not based on sufficiently reliable principles and methodologies under Daubert/Kumho and fully excluded the expert's testimony. *United States v. Lewis*, 220 F. Supp. 2d 548 (S.D.W. Va. 2002); *United States v. Brewer*, No. 01 CR 892, 2002 U.S. Dist. LEXIS 6689 (N.D.Ill 2002); *United States v. Fujii*, 152 F. Supp. 2d 939 (N.D.Ill. 2000). Three courts reached a middle position, permitting the FDE to testify as to particular similarities and dissimilarities between the documents, but excluding the ultimate opinion as to authorship. *United States v. Rutherford*, 104 F. Supp. 2d 1190 (D. Neb. 2000); *United States v. Santillan, No. CR-96-40169*, 1999 U.S. Dist. LEXIS 21611 (N.D.Cal. 1999); *United States v. Hines*, 55 F. Supp. 2d 62 (D.Mass. 1999).

[168] Harrison, Burkes, and Seiger, "Handwriting Examination: Meeting the Challenges of Science and the Law."

[169] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 27..

The first principle implies that handwriting is unique to an individual, which has motivated a body of research on the individualization of handwriting.[170] As outlined in section 1.1, the conventional belief in individuality stemmed from early writings of Osborn[171] and continues among FDEs today.[172] However, FDE decision making does not depend on the concept of uniqueness[173] but rather the rarity of the features. Uniqueness lies at the very extreme of the spectrum from rare to common features; FDEs do not need to claim that an exemplar is unique to claim writership. Because every instance of handwriting is unique in that it is characterized by a distinct set of habitual features, claiming uniqueness is not a useful indicator of writership.

Early practitioners of handwriting examination relied on established statistical rules to support the principle of individuality. For example, Osborn[174] applied the Newcomb rule[175] of probability to demonstrate how combinations of similar writing habits from two samples could occur with a frequency derived by multiplying together the respective ratios of frequencies of occurrence of each of the habits. Unfortunately, Osborn did not consider the dependencies between the variables in Newcomb's rule. Nevertheless, the rule and Osborn's interpretation were accepted as the principle of identification[176] in handwriting examination. Huber stated that[177]

> [w]hen any two items possess a combination of similar and independent characteristics, corresponding in relationship to one another, of such number and significance as to preclude the possibility of coincidental occurrence, without inexplicable disparities, it may be concluded that they are the same in nature or are related to a common source.

A more contemporary view of individuality refers to a given population of writers studied with a given comparison methodology. In this view, individuality is defined with respect to the probability of observing writing profiles of two individuals that are indistinguishable using the

---

[170] M. Beacom, "A Study of Handwriting by Twins and Other Persons of Multiple Births," *Journal of Forensic Sciences* 5, no. 1 (1960); D. Boot, "An Investigation into the Degree of Similarity in the Handwriting of Identical and Fraternal Twins in New Zealand," *Journal of the American Society of Questioned Document Examiners* 1 (1998); S. Lines and F. E. Franck, "Triplet and Sibling Handwriting Study to Determine Degree of Individuality and Natural Variation," *Journal of the American Society of Questioned Document Examiners* 6 (2003); D. J. Gamble, "The Handwriting of Identical Twins," *Canadian Society of Forensic Science Journal* 13, no. 1 (1980), https://doi.org/10.1080/00085030.1980.10757337; Sargur N. Srihari et al., "Individuality of Handwriting," *Journal of Forensic Sciences* 47, no. 4 (2002), https://doi.org/10.1520/jfs15447j; S. Srihari, C. Huang, and H. Srinivasan, "On the Discriminability of the Handwriting of Twins," *Journal of Forensic Sciences* 53, no. 2 (Mar 2008), https://doi.org/10.1111/j.1556-4029.2008.00682.x.

[171] A. S. Osborn, *Questioned Documents*, Second ed. (Albany, NY: Boyd Printing Company, 1929).

[172] The assumption of uniqueness in forensic identification sciences has been attacked as "metaphysical" (J. Koehler and M. J. Saks, *Individualization Claims in Forensic Science: Still Unwarranted* (2010), https://scholarlycommons.law.northwestern.edu/facultyworkingpapers/27. But see D. H. Kaye, "Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies," *Brooklyn Law Review* 75 (2010), https://elibrary.law.psu.edu/cgi/viewcontent.cgi?article=1015&context=fac_works.)

[173] See discussion in M. Page, J. Taylor, and M. Blenkin, "Uniqueness in the Forensic Identification Sciences—Fact or Fiction?," *Forensic Science International* 206, no. 1-3 (Mar 20 2011), https://doi.org/10.1016/j.forsciint.2010.08.004.

[174] Osborn, *Questioned Documents*, 226.

[175] Osborn, *Questioned Documents,* 266 provides a definition of the Newcomb rule as "The probability of occurrence together of all the events is equal to the continued product of the probabilities of all the separate events."

[176] SWGDOC defines identification ("definite conclusion of identity") as "the highest degree of confidence expressed by document examiners in handwriting comparisons. The examiner has no reservations whatever, and although prohibited from using the word 'fact,' the examiner is certain, based on evidence contained in the handwriting, that the writer of the known material actually wrote the writing in question. Examples—It has been concluded that John Doe wrote the questioned material, or it is my opinion [or conclusion] that John Doe of the known material wrote the questioned material." See Scientific Working Group for Forensic Document Examination (SWGDOC). *SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners*.

[177] R. A. Huber, "Expert Witnesses," *Criminal Law Quarterly* 2, no. 3 (1959).

specified comparison method.[178] The greater the degree of individuality in the population, the less likely it is that the writing profiles of two individuals would be observed as indistinguishable.[179]

Uniqueness and individualization in forensic science no longer correspond to the conventional, strict interpretation of these terms[180] and can lead to an exaggeration of the strength of the evidence. Indeed, empirical research and statistical reasoning do not support source attribution to the exclusion of all others. In practice, FDEs often (but not always) explain in reports and testimony that an identification that excludes all others cannot be proven.

Thus, the Working Group makes the following recommendation:

Recommendation 2.3: Forensic document examiners must not report or testify, directly or by implication, that questioned handwriting has been written by an individual (to the exclusion of all others).

2.2.1.1. Moving Away from Conventional Principles in Forensic Handwriting Examination

Although conventional principles underlying handwriting examination like feature comparison remain relevant, appreciation of the source and range of natural variation both between and within individuals is more important. The causes of intra- and inter-writer variation, and the arguments for why intra-writer variation is smaller than inter-writer variation, have deep roots in motor control theory.

Motor control theory is based on neurobiological principles. The theory treats the handwritten stroke to be the base unit. The temporal and geometric properties of handwriting strokes are programmed, sequenced, and executed by the central nervous system. Over time, an individual learns or habituates complex sequences of motor commands, reducing the demands placed on memory and motor systems during natural writing.[181] As the complex motor sequences of handwriting become habituated over time, the feature variability exhibited by individuals decreases within an individual writer while the flexibility to adapt to changing spatial or physical constraints increases. These properties enable several predictions about writership variability, including the prediction that certain features of handwriting remain invariant throughout changes in writing surface, orientation, or whether the individual wrote with the dominant or non-dominant hand. This is referred to as the principle of motor equivalence,[182] defined by Lashley[183] as observations of variable means to invariant ends. This and other aspects of motor

[178] Srihari et al., "Individuality of Handwriting."

[179] C. P. Saunders, L. J. Davis, and J. Buscaglia, "Using Automated Comparisons to Quantify Handwriting Individuality," *Journal of Forensic Sciences* 56, no. 3 (May 2011), https://doi.org/10.1111/j.1556-4029.2011.01713.x.

[180] See D. H. Kaye et al., *The New Wigmore: A treatise on Evidence: Expert Evidence* (Austin, TX: Aspen Publishers, 2011).“General uniqueness” means that every element of a set is distinguishable from every other element. “Special uniqueness” means that a particular element is distinguishable from all others even if not all of the remaining elements are each distinguishable. D. H. Kaye, "Identification, Individualization and Uniqueness: What's the Difference?," *Law, Probability and Risk* 8, no. 2 (2009), https://doi.org/10.1093/lpr/mgp018.

[181] M. P. Caligiuri and L. A. Mohammed, "Chapter 3," in *The Neuroscience of Handwriting* (Boca Raton: CRC Press, 2012).

[182] Alan M. Wing, "Motor Control: Mechanisms of Motor Equivalence in Handwriting," *Current Biology* 10, no. 6 (2000), https://doi.org/10.1016/s0960-9822(00)00375-4.

[183] K. S. Lashley, "Mass Action in Cerebral Function," *Science* 73, no. 1888 (Mar 6 1931), https://doi.org/10.1126/science.73.1888.245.

control theory (e.g., complexity theory[184]) when applied to handwriting have the potential to shift the foundation of handwriting examination from the assumptions of individualization (i.e., the conventional Osbornian approach) to an empirical neurobiological approach that allows for hypothesis generation, predictions about handwriting variability, and research of questions relevant to the handwriting examination.

Among the empirically tested motor control hypotheses, motor equivalence stands out for its relevance to handwriting examination. Motor equivalence[185] makes two important predictions to handwriting examination. The first is the existence of a motor program as a theoretical memory structure capable of transforming an abstract code into an action sequence. Regarding handwriting, the timing and sequence of pen strokes produced to form letters and words or a signature are stored in a flexible, generalized motor program available to the writer as a single action sequence. Such a memory structure might contain a fixed set of commands timed in such a way that movement parameters like torque, trajectory, speed, and distance may be reliably repeated. Motor equivalence also predicts that these action sequences can adapt to environmental or internal alterations such that the handwriting control sequences can be faithfully executed despite differences in writing surface, writing instrument, or special constraints.[186]

The presence of inter- and intra-writer variation in forensic handwriting examination does not imply that evidence of marked feature variation should lead to an opinion that questioned handwriting samples may be from different writers. Hilton[187] and other authors[188] have addressed the issue of the relative importance of inter-writer variation in forensic handwriting examinations. These authors state that a difference that is fundamental in nature is compelling and a sufficient basis for "non-identity." Harrison has asserted that two samples of handwriting "cannot be considered to be of common authorship if they display but a single consistent dissimilarity in any feature which is fundamental to the structure of the handwriting, and whose presence is not capable of reasonable explanation."[189] Some FDEs take this to mean that even a single fundamental difference is grounds for the elimination of the subject writer as having prepared the entry in question. However, to establish that a dissimilarity is a true difference, the FDE must be able to reasonably exclude any potential distortion because of any internal or external factors. In addition, the FDE must determine that the submitted known specimens fully reflect the specimen writer's entire range of variation at the specific time of the questioned writing's execution and under a plethora of circumstances.

The exclusion of all these possible effects would be a complex and daunting task even under ideal circumstances. An FDE's report that eliminates a writer as the source of a questioned entry

[184] J. J. Brault and R. Plamondon, "A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery," *IEEE Transactions on Systems, Man, and Cybernetics* 23, no. 2 (1993), https://doi.org/10.1109/21.229453; B. Found et al., "Statistical Modelling of Experts' Perceptions of the Ease of Signature Simulation," *Journal of Forensic Document Examination* 11 (1998).

[185] Caligiuri and Mohammed, "Chapter 3."

[186] Wing, "Motor Control: Mechanisms of Motor Equivalence in Handwriting."

[187] O. Hilton, *Scientific Examination of Questioned Documents*, Revised ed. (New York: Elsevier North Holland, Inc., 1982), 10.

[188] Kelly and Lindblom, *Scientific Examination of Questioned Documents*, 63; Osborn, *Questioned Documents*, 262; W. R. Harrison, *Suspect Documents: Their Scientific Examination* (London: Sweet & Maxwell Limited, 1958), 343.

[189] Harrison, *Suspect Documents: Their Scientific Examination*, 343.

based solely on one fundamental difference should be viewed with skepticism. Hilton[190] and Harrison[191] both noted that multiple characteristic differences—not just one—will be found.

Brault and Plamondon[192] developed an imitation (forgery) difficulty coefficient based on a formula that models the complex processes involving perception, memorization, and muscle coordination that the imitator employs to execute a forgery. Line length, stroke duration, and angularity of turning points were included in the formula. The higher the difficulty coefficient, the larger the variation in one person's genuine signature can be and, therefore, the lower the threshold for a new signature to be accepted as valid. Similarly, Found et al.[193] and Alewijnse et al.[194] analyzed which factors make a signature difficult to simulate. They observed that the number of turning points and line intersections or retraces best explain the FDE's assessment of signature complexity. By considering the neuromotor factors underlying signature production, FDEs can more accurately predict the presence of feature sets or patterns that should characterize genuine and simulated or disguised signatures.

2.2.1.2. Reliability of the Method of Analysis

Several guidance documents prepared for the forensic community address the validity and reliability of analysis methods. These documents are listed below.

- 2009 NRC of the National Academy of Sciences (NAS) report on strengthening forensic science in the United States[195]
- ENFSI Best Practice Manual for the Forensic Examination of Handwriting[196]
- Latent Print Examination and Human Factors report (Latent Print report)[197]
- Fundamentals of Probability and Statistical Evidence in Criminal Proceedings, published by the Royal Statistical Society[198]
- 2016 President's Council of Advisors on Science and Technology (PCAST) report on ensuring scientific validity of feature comparison methods[199]

---

[190] Hilton, *Scientific Examination of Questioned Documents*, 10.

[191] Harrison, *Suspect Documents: Their Scientific Examination*, 345.

[192] Brault and Plamondon, "A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery."

[193] Found et al., "Statistical Modelling of Experts' Perceptions of the Ease of Signature Simulation."

[194] L. C. Alewijnse, C. E. van den Heuvel, and R. D. Stoel, "Analysis of Signature Complexity," *Journal of Forensic Document Examination* 21 (2011).

[195] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*.

[196] *Best Practice Manual for the Forensic Examination of Handwriting,* ENFSI-BPM-FHX-01, (European Network of Forensic Science Institutes (ENFSI), June, 2018).

[197] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*.

[198] C. Aitken, P. Roberts, and G. Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*, Royal Statistical Society (2010), https://www.researchgate.net/publication/259088224_Fundamentals_of_Probability_and_Statistical_Evidence_in_Criminal_Proceedings_Guidance_for_Judges_Lawyers_Forensic_Scientists_and_Expert_Witnesses.

[199] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, President's Council of Advisors on Science and Technology (Washington, DC, 2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

We note that definitions relating to validity and reliability may differ depending on the paradigm and context in which they are used. Box 2.2 provides an explanation of these terms in the context of forensic handwriting examination and within this report.

Box 2.2: Reliability and validity in the context of forensic handwriting examination

Reliability: To what degree single or multiple FDEs reach the same answer under specified tasks and constant conditions. Reliability is related to the degree of random error of the instrument/method, which can include the FDE. The smaller the amount of random error, the more reliable the instrument/method, and vice versa. Two ways to assess reliability are repeatability and reproducibility.[200]

> **Repeatability:** A measure of reliability using the same FDE and the same instrument/method under exactly the same conditions to arrive at the same conclusion or result.
>
> **Reproducibility:** A measure of reliability using different FDEs and/or differing conditions with the same measurement instrument/method to arrive at the same conclusion or result.

Validity: To what degree single or multiple FDEs reach the correct answer under specified tasks and constant conditions. A test is valid if it measures what it is supposed to measure.[201] A measure can be reliable and not valid, but not vice versa. In other words, reliability is necessary but not sufficient for validity, and, if a measurement instrument/method is valid, it is also reliable.

Accuracy: Similar to validity in that it relates to correctness of a result (i.e., closeness of measurements/outcomes to the true value).

Systematic error: A component of error whereby replicate measurements remain constant or vary in a predictable way—for example an uncalibrated instrument would produce a constant systematic error.[202]

Random error: A component of error whereby replicate measurements vary in an unpredictable way. Sources of random error are usually unexplained and therefore difficult to control.[203]

The NRC report on strengthening forensic science in the United States cautions that "the interpretation of forensic science is not always based on scientific studies to determine its

[200] For application of the concepts discussed under reliability to forensic science, see Ulery et al., "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners."

[201] See Borsboom, Mellenbergh, and van Heerden, "The Concept of Validity."

[202] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

[203] Joint Committee for Guides in Metrology (JCGM), "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)."

validity."[204] The report pointed to the general requirements under ISO/International Electrotechnical Commission (IEC) 17025:2005[205] for competence testing and laboratory calibration as a source of well-established approaches to validating a method. These include (1) calibration using a standard reference, (2) ensuring agreement between two uncorrelated methods in reaching the same result, (3) inter-laboratory comparisons, (4) assessing factors that could influence a result, and (5) assessment of the uncertainty of the result based on knowledge of the scientific and theoretical principles underlying the method. Furthermore, the NRC noted that publication in peer-reviewed journals is also an important component of the validation process, because it enables experts to review research critically and attempt to replicate results.

The ENFSI approach to process validation broadens the more conventional criteria by considering FDE competence and quality control as bare minimums to establish the validity of an examination procedure. The ENFSI guidance document includes the following minimum requirements for a forensic examination procedure to be considered valid:[206]

- There is an agreed requirement for the technique or procedure.
- The critical aspects of the technique or procedure have been identified and the limitations defined.
- The methods, materials, and equipment used have been demonstrated to be fit for purpose in meeting the requirement.
- There are appropriate QC and QA procedures in place for monitoring performance.
- The technique or procedure is fully documented.
- The results obtained are reliable and reproducible.
- The technique or procedure has been subjected to an independent assessment and, where novel, peer review.
- The individuals using the technique or procedure have demonstrated that they have been trained and that they are competent.

With its focus on human factors, the Working Group's viewpoint more closely aligns with the latent print Expert Working Group,[207] which discussed error rates, and in discussing validation, focused on whether "measurements, judgments, and decisions being made are appropriate for their common uses."[208] This reference to common use is in agreement with the ENFSI requirement that a procedure needs to be appropriate for purpose to be deemed valid. As characterized in the Latent Print report, validity is a relative term. In other words, demonstrating that comparison procedures may be valid to evaluate the evidence given one set of propositions

[204] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 8.

[205] *General Requirements for the Competence of Testing and Calibration Laboratories,* ISO/IEC 17025:2005, (Geneva, Switzerland: International Organization for Standardization (ISO), 2005). Section 5.4.5 2 (Note 2).

[206] European Network of Forensic Science Institutes (ENFSI). *Best Practice Manual for the Forensic Examination of Handwriting*, 8.

[207] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 74.

[208] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 75.

does not imply that the same procedures are valid for evaluating the evidence given another set of propositions. For example, the extent to which feature comparisons are considered valid will depend on whether the methods are designed to serve that specific purpose (e.g., comparing or measuring attributes of genuine versus simulated signatures might not be valid for hand-printed material).

Inattention to method validation may lead to errors like data misrepresentation, inadequate method selection, and unreliable conclusions about evidentiary strength.

2.2.2. Reliability and Validity in Handwriting Examination

The terms validity and reliability are used differently in legal discourse than in science.[209] In science, reliability often refers to the output consistency of a test or measuring device. A scale, for example, is reliable if it reports the same weight for the same object time and again. Unreliability can be measured by how much variation exists among repeated outputs to a given input or among different measuring devices to a given input. The measurement device may not be accurate—it may always report a weight that is too high or too low—but the reliable scale always reports the same weight for the same object. Its errors, if any, are systematic.

As stated in the NRC report: "[a] key task… for the analyst applying a scientific method is to conduct a particular analysis to identify as many sources of error as possible, to control or eliminate as many as possible, and to estimate the magnitude of remaining errors so that the conclusions drawn from the study are valid."[210] In other words, there will always be an element of uncertainty in every measurement. The uncertainty stems from the fact that the true value of the measurement is never exactly known. In handwriting comparisons, potential sources of systematic error include the FDE and the workflow process/method (see chapter 1), each of which can be minimized with an understanding of the contribution these factors play in validating an evaluative process.

Two different aspects of reliability should be considered: intra-examiner (i.e., within the observer) and inter-examiner (i.e., between observers). Variability in intra-examiner judgements should be small. That is, the same evaluator should rate essentially identical cases in similar ways. Variability in inter-examiner judgements should be small. That is, different evaluators should rate the same cases in essentially the same way.

Without agreement between independent observers who are able to reproduce procedures or the ability to use tools and procedures that yield consistent measurements/outcomes, researchers cannot satisfactorily draw conclusions, formulate theories, or make claims about the generalizability of their observations. While validity is concerned with the degree of success at measuring what the research set out to measure, reliability is concerned with the consistency of the actual measuring instrument or procedure.

[209] In legal discourse, "reliability" often means the plausibility or credibility of an assertion, which fuses the scientific concepts of validity and reliability. See, for example, Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S. Proposed testimony must be supported by appropriate validation—i.e., "good grounds," based on what is known. In short, the requirement that an expert's testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability.

[210] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*.

Reliability and validity have a nested relationship. Reliability is a necessary but not sufficient condition of validity.[211] As noted, a reliable process can be invalid if it consistently measures something other than the outcome of interest it is being used to measure. An unreliable process undermines validity.

In practice, the term reliability is used to mean the consistency of a measure or interpretation. As noted in box 2.2, to establish the reliability of measurement (or a process), one must have repeatability (intra-examiner consistency) and reproducibility (inter-examiner consistency). To be valid, a measure (or interpretation) must have not only inter- and intra-examiner consistency, but it must also measure what it intends to measure. In other words, for an instrument (or FDE in the case of handwriting) to yield consistent results or observations, relevant systematic error (e.g., bias) must be minimized in either the instrument or the interpretation of the data. As noted in the Latent Print report, "[e]stablishing reproducibility, therefore, is a part of the process of validating measurements, but concordance between the two examiners is a flawed measure even of reproducibility if the verifying examiner's judgments are influenced by knowledge of the first examiner's opinion."[212] Although the criteria proposed in the PCAST report[213] underscore the importance of reproducibility, repeatability, and accuracy, the possibility remains that a process derived from flawed scientific principles or constructs, if reproducible, might be mistaken as valid.

To estimate repeatability and reproducibility of judgements in handwriting examination, studies should compare the performance within and between FDEs in their judgements on the same samples of handwriting against ground truth. If the same FDE repeatedly reaches the same conclusions (whether right or wrong) on the same set of handwriting tasks in examinations separated by sufficient time, intra-examiner reliability (for the test samples) is high. Similarly, if multiple FDEs independently performing the same handwriting tasks reach the same conclusions, inter-examiner reliability (for the test samples) is high. Although the PCAST report[214] recommends imposing the requirement of reproducibility testing by multiple independent FDEs, it is not self-evident that *Daubert*[215] makes the same requirement. The Working Group's view is that multiple independent laboratories should collaborate to address the problem of repeatability and reproducibility using the same materials and methods.

[211] See J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, 3rd ed. (New York: McGraw-Hill Publishing Co., 1994), 111; R. E. Kirk, *Experimental Design*, 2nd ed. (Belmont, CA: Wadsworth, Inc, 1982); E. G. Carmines and R. A. Zeller, *Reliability and Validity Assessment*, Sage Publications (London, 1979), https://academic.csuohio.edu/kneuendorf/c63111/C&ZSage.pdf.

[212] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 34.

[213] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 106.

[214] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*.

[215] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.

In addition to numerous studies of cognitive bias,[216] a small but growing number of studies of forensic examiners have investigated whether biasing information produces changes in expert judgements. In a meta-analysis of small-scale studies of fingerprint experts, Dror and Rosenthal[217] concluded that experts were neither reliable (when presented a second time with historical cases they had previously reviewed) nor unbiased (when the context was manipulated to examine whether extraneous information might bias the expert).

Upon comparing handwriting samples, FDEs gauge the strength of their belief on scales ranging from the three-point scale (same source, inconclusive, or different source) to the more elaborate SWGDOC nine-point classification scheme (see table 1.4). The intra-examiner reliability of these scales has not been subjected to rigorous empirical study. In designing such studies, investigators should include random repeats of sample pairs to assess the consistency of FDE judgement.

Factors underlying the reliability of the process are likely to differ from those contributing to the reliability of the decisions rendered. Studies are needed to test whether steps along the process map in figure 1.1 are comprehensively reflective of actual casework and if different FDEs using the same process reach the same conclusions. It is unclear whether the process needs to be strictly followed to attain high levels of inter- and intra-examiner reliability and which elements of the process, if any, contribute to FDE inconsistency.

Empirical studies that can speak to the reliability of outputs are typically referred to as "black box" tests. For black box tests, the methods used by the test subjects are unknown. Different FDEs performing subjective feature comparison methods like handwriting examination may detect or focus on different features, attach differing levels of importance to the same features, and have different criteria altogether for reaching a conclusion. However, the procedures for decision making at these stages are generally not objectively specified, so the overall procedure must be treated as a black box inside the FDE's head.[218]

---

[216] Dror and Charlton, "Why Experts Make Errors."; Dror, Charlton, and Peron, "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications."; Itiel E. Dror and G. Hampikian, "Subjectivity and Bias in Forensic DNA Mixture Interpretation," *Science & Justice: Journal of the Forensic Science Society* 51, no. 4 (Dec 2011), https://doi.org/10.1016/j.scijus.2011.08.004; Dror et al., "Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison."; Fraser-Mackenzie, Dror, and Wertheim, "Cognitive and Contextual Influences in Determination of Latent Fingerprint Suitability for Identification Judgments."; L. J. Hall and E. Player, "Will the Introduction of an Emotional Context Affect Fingerprint Analysis and Decision-Making?," *Forensic Science International* 181, no. 1-3 (Oct 25 2008), https://doi.org/10.1016/j.forsciint.2008.08.008; Kerstholt et al., "Does Suggestive Information Cause a Confirmation Bias in Bullet Comparisons?."; Nakhaeizadeh, Dror, and Morgan, "Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias."; Osborne et al., "Does Contextual Information Bias Bitemark Comparisons?."; Osborne et al., "Bloodstain Pattern Classification: Accuracy, Effect of Contextual Information and the Role of Analyst Characteristics."; M. Page, J. Taylor, and M. Blenkin, "Forensic Identification Science Evidence Since Daubert: Part II—Judicial Reasoning in Decisions to Exclude Forensic Identification Evidence on Grounds of Reliability," *J Forensic Sci* 56, no. 4 (Jul 2011), https://doi.org/10.1111/j.1556-4029.2011.01776.x; Risinger et al., "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion."; B. Schiffer and C. Champod, "The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation," *Forensic Science International* 167, no. 2-3 (Apr 11 2007), https://doi.org/10.1016/j.forsciint.2006.06.036; W. C. Thompson, "Painting the Target Around the Matching Profile: the Texas Sharpshooter Fallacy in Forensic DNA Interpretation," *Law, Probability and Risk* 8, no. 3 (2009), https://doi.org/10.1093/lpr/mgp013; Miller, "Bias Among Forensic Document Examiners: A Need for Procedural Change."; Langenburg, Champod, and Wertheim, "Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons."; S. Nakhaeizadeh, I. E. Dror, and R. Morgan, "The Emergence of Cognitive Bias in Forensic Science and Criminal Investigations," *British Journal of American Legal Studies* 4 (2015).

[217] Dror and Rosenthal, "Meta-Analytically Quantifying the Reliability and Biasability of Forensic Experts."

[218] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 5.

Black box studies require many FDEs to render opinions about many independent comparisons (typically, involving questioned samples and one or more known samples) so that error rates can be determined.[219] However, the utility of a global error rate as determined by a black box study is questionable, because the rate is only relevant to the conditions within that particular test, and it does not necessarily speak to the source or cause of the error.[220]

"White box" tests, alternatively, are designed to help understand the factors (e.g., quality and quantity of questioned material) that affect FDEs' decisions. These factors are made known—meaning they are also useful in determining sources of error. In these tests, samples represent the variable of interest and may require application of only a portion of the feature comparison method.

Results of black box and white box tests in handwriting examination may lead to a refinement of the process map and, ultimately, improved reliability.[221] The Hierarchy of Expert Performance (HEP) may assist in designing such studies systematically.[222] HEP can be used to quantify expert performance by systematically examining reliability and biasability between and within experts and by separating observations from conclusions. Evaluating expert performance within HEP facilitates the identification of strengths and weaknesses in expert performance and enables the comparison of experts across domains. HEP may also provide theoretical and applied insights into expertise.

Therefore, the Working Group makes the following recommendation.

Recommendation 2.4: Forensic document examiners should collaborate with researchers to design and participate in "black box" and "white box" studies.

2.3. Interpreting Handwriting Evidence

2.3.1. Feature Selection and Interpretation

Steps 300 and 700 of the process map (see figure 1.1) direct FDEs to select features from questioned and known handwriting exemplars that they identify as important to the examination. Feature selection often depends on the presence of unusual or potentially discriminating characteristics. Although selecting features for examination is largely subjective and therefore vulnerable to contextual bias (see section 2.1), it is important to capture discriminating features to ensure a more accurate interpretation.

Currently, there are four basic approaches to feature selection:

---

[219] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 5–6.

[220] See T. R. Hunt, "Scientific Validity and Error Rates: A Short Response to the PCAST Report," *Fordham Law Review Online* 86, no. 14 (2017): 35, https://ir.lawnet.fordham.edu/flro/vol86/iss1/14.

[221] President's Council of Advisors on Science and Technology (PCAST), *An Addendum to the PCAST Report on Forensic Science in Criminal Courts* (January 6, 2017 2017), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf.

[222] Itiel E. Dror, "A Hierarchy of Expert Performance," *Journal of Applied Research in Memory and Cognition* 5, no. 2 (2016), https://doi.org/10.1016/j.jarmac.2016.03.001; Itiel E. Dror and Daniel C. Murrie, "A Hierarchy of Expert Performance Applied to Forensic Psychological Assessments," *Psychology, Public Policy, and Law* 24, no. 1 (2018), https://doi.org/10.1037/law0000140.

1. Use a generally accepted, predefined set of features and their relative frequency of occurrence in a specified population.[223]
2. Use the questioned documents to suggest the features of interest before a side-by-side comparison.
3. Use the known documents to suggest the features of interest before a side-by-side comparison.
4. Use both the questioned and reference writings side-by-side during the feature selection process.

A comprehensive, predefined set of features indicating their rarity within a representative population does not currently exist in a way that is easy for FDEs to apply in all cases. Research[224] has been performed to begin the process of developing a predefined set of features. If that set were available, it may contribute to a more objective process, less affected by potential FDE bias than other approaches. Using the questioned document to suggest the features of interest is not as objective as a predefined feature set. However, it might be less susceptible to bias than using the known writing to suggest features for comparison or a side-by-side comparison to select features, which may increase the risk of bias. See section 2.1 for further discussion on such bias.

In some fields, probability models and data on the distribution of features in relevant populations permit forensic scientists to calculate the strength of evidence. The best example is forensic DNA analysis. Many human population samples exist for estimating how often variants of a particular genetic marker are present in the population, and a well-defined model for combining them into a profile frequency is available, as well as data on measurement uncertainty. In other fields, analogous data and models either do not yet exist or have been developed but are still being validated. FDEs currently have limited data on how often particular features occur in nature. Nevertheless, they can draw on existing information, existing databases, and newly constructed databases,[225] along with their general knowledge and experience, to judge how strongly the observed features in the questioned and known writings (i.e., the evidence) support the propositions of interest in a particular case.[226]

At various points in the handwriting examination process, an FDE decides whether the exemplar is of value for numerous purposes and makes decisions regarding the sufficiency or suitability for comparison, including

1. Feature sufficiency. An FDE decides whether there is an adequate amount of information available for comparison.

[223] Huber and Headrick, *Handwriting Identification: Facts and Fundamentals*, 136–38.

[224] Johnson et al., "Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics."

[225] Johnson et al., "Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics."

[226] These propositions are often denoted as the “prosecution proposition” versus the “defense proposition,” but they can be formulated before any prosecution commences.

2. Feature weighting. An FDE assigns a value and significance to individual features and their configuration and assesses the overall strength of their synthesis. Interpretative errors can occur when an FDE excludes relevant features or fails to assign appropriate weight to the feature.

3. Feature discrepancy. An FDE interprets the significance of observed divergences between handwriting exemplars to determine whether the feature differences are indicative of different sources or indicative of a common origin. To make this interpretation, the FDE must have knowledge of the frequency of occurrence of the identified features within the relevant population. Without objective datasets, this interpretation is informed by the FDE's knowledge and experience.

2.3.2. Handwriting Comparison Approach and Evaluation

Chapter 1 describes the conventional process by which an FDE compares questioned and known samples of handwriting to address the proposition that the samples originated from the same writer. In this conventional approach (also referred to as the classical approach or two-stage approach[227]), the FDE seeks to reach a conclusion from the perspective of the proposition, like the signature was produced by the person of that name or the threatening letter was (not) written by the suspect. For brevity, such propositions are denoted as H_1 (and H_2) and the putative writer as W_1. Conventionally, an FDE might opine with a high degree of certainty that the writer is individualized, based on the classical premise that no one else in the relevant population could have signed the name or written the words on the questioned document.

In a variant of this approach, the FDE will first decide whether the suspect could have written the questioned document based on the similarities and dissimilarities observed between the questioned document and the known writing samples. If the suspect writer cannot be excluded as the writer of the questioned document, the FDE then considers the rate at which alternative writers cannot be excluded as the source of the questioned document. This rate can be referred to as the "coincidence probability."[228] If the suspect cannot be excluded and the coincidence probability is sufficiently low, then the evidence is in favor of H_1; the larger the coincidence probability, the weaker the evidence becomes. Some literature on forensic statistics debates the reasonableness of the coincidence probability,[229] which in a handwriting examination context corresponds to the rate at which alternative sources match the questioned document. An additional variant is added by mapping these coincidence probabilities to a reporting scale with a set of ordered categories like "true," "false," or "inconclusive," perhaps adding terms like

[227] J. B. Parker, "A Statistical Treatment of Identification Problems," *Journal of the Forensic Science Society* 6, no. 1 (1966); I. W. V. Evett, "The Interpretation of Refractive Index Measurements," *Forensic Science* 9 (1977), https://doi.org/10.1016/0300-9432(77)90093-0.

[228] See J. M. Curran, T. N. Hicks, and J. S. Buckleton, *Forensic Interpretation of Glass Evidence* (Boca Raton: CRC Press—Taylor & Francis Group, 2000); J. Buckleton, C. M. Triggs, and S. J. Walsh, *Forensic DNA Evidence Interpretation* (Boca Raton, FL: CRC Press, 2005); I. W. Evett and J. A. Lambert, "The Interpretation of Refractive Index Measurements. III," *Forensic Science International* 20, no. 3 (1982), https://doi.org/10.1016/0379-0738(82)90123-2; I. E. Evett, "Interpretation: a Personal Odyssey," in *The Use Of Statistics In Forensic Science*, ed. C. G. G. Aitken and D. A. Stoney (London: CRC Press, 1991); D. A. Stoney, "Evaluation of Associative Evidence: Choosing the Relevant Question," *Journal of the Forensic Science Society* 24, no. 5 (1984), https://doi.org/10.1016/s0015-7368(84)72326-7.

[229] Curran, Hicks, and Buckleton, *Forensic Interpretation of Glass Evidence*; Stoney, "Evaluation of Associative Evidence: Choosing the Relevant Question."

"strong probability," "probable," and "indications."[230] Even though the coincidence probability is defined as a frequentist probability, it is typically estimated in a subjective manner based on the FDE's experience and then mapped to a conclusion scale.

All these types of evaluative statements share a common thread. They presuppose that the FDE's task is to give some opinion in support of any proposition, here referred to as H_1 (if the samples are adequate to perform an examination). However, the usefulness and appropriateness of this conventional interpretative framework have been questioned.[231] In particular, one can question the premise that the expert should come to any decision (qualified or otherwise) about H_1.[232] Although expert opinions about matters that a judge or jury must ultimately resolve are generally permissible, they are not required by any rule of law or scientific principle.[233] The expert need not proffer an opinion about H_1—or be compelled to do so—to contribute scientific information to the resolution of a case.[234]

For example, although some courts have excluded the conventional conclusion-oriented testimony, there have been some instances where a features-only testimony has been permitted, and the expert is limited to a description of the relevant features of the samples. The underlying idea is that the expert has ample knowledge to point out salient features, including "things that the jury might not see on its own."[235] The jurors then "can use their own powers of observation and comparison"[236] "to make the ultimate finding of identity or non-identity."[237] A major issue with this features-only approach is that it forces jurors to interpret and perform inferential tasks themselves—a task they have neither trained in nor practiced. By confining the expert interpretation to feature identification and precluding expert inferences from these observations, jurors may overestimate (or underestimate) the probative value of the handwriting evidence, erroneously giving more (or less) weight to some similarities or differences than others.

There is increasing consensus that expert testimony would most effectively assist the court or jury to reach its conclusion about H_1 if it is based on the extent to which the findings (i.e., the degree of correspondence between the samples) supports H_1 relative to one or more alternative propositions. The important development of this paradigm is the reporting of the relative support for one proposition over another proposition, without addressing the probability of the propositions themselves (see the conclusion scales in figure 3.1 for details). This mode of

---

[230] Scientific Working Group for Forensic Document Examination (SWGDOC). *SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners*.

[231] For example, D. J. Balding, *Weight-of-Evidence for Forensic DNA Profiles* (Hoboken: John Wiley & Sons, 2005).

[232] W. A. Wagenaar, *Identifying Ivan: A Case Study in Legal Psychology* (London: Harvester/Wheatsheaf, 1988).

[233] Kaye et al., *The New Wigmore: A treatise on Evidence: Expert Evidence*; B. Robertson, G. A. Vignaux, and C. E. H. Berger, *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, 2nd ed. (Chichester: Wiley, 2016).

[234] G. Jackson, C. Aitken, and Roberts P., *Case Assessment and Interpretation of Expert Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*, Royal Statistical Society (London, 2014), https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.723.2831&rep=rep1&type=pdf.

[235] *United States v. Hidalgo*, 229 F. Supp. 2d 961, 968 (D. Ariz. 2002).

[236] *State v. Reid*, 757 A.2d 482, 487 (Conn. 2000). discussing features-only testimony about a microscopic hair comparison.

[237] United States v. Hidalgo, F. Supp. 2d. explains that "[w]hile the failure of proof of the uniqueness principle would preclude him from rendering an opinion of identity, he could, based upon his experience and training, testify to the mechanics and characteristics of handwriting, his methodology, and his comparisons of similarities and dissimilarities between the defendants known writings and those of the questioned documents. https://law.justia.com/cases/federal/district-courts/FSupp2/229/961/2396837/

evaluation and reporting, described in papers and books[238] for more than 50 years, is called the "Bayesian approach" or the "likelihood ratio approach" and has been adopted by a small number of forensic laboratories around the world.[239] It diverges from the conventional mode of giving the fact finder some degree of confidence about a categorical source attribution. It asks the expert to limit evaluative conclusions to the degree of support that the evidence provides for H_1 compared with the alternative H_2. This approach makes explicit that the evaluation of forensic science evidence is always conducted in a framework of task-relevant background information and is always relative to specified and explicit competing propositions for how the evidence has arisen. Different framework information or propositions will result in a different evaluation and, consequently, may lead to a different conclusion.

In the likelihood ratio approach, one has to find a proper way to measure the support that the findings have for each proposition (see box 2.3). Many advocate[240] that probability is the best candidate for forensic identification of source problems, although some researchers have criticized[241] this approach.

[238] Including Aitken, Roberts, and Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*; Association of Forensic Science Providers, "Standards for the Formulation of Evaluative Forensic Science Expert Opinion," *Science & Justice* 49, no. 3 (2009); J. S. Buckleton, C. M. Triggs, and C. Champod, "An Extended Likelihood Ratio Framework for Interpreting Evidence," *Science & Justice* 46, no. 2 (2006), https://doi.org/10.1016/s1355-0306(06)71577-5; European Network of Forensic Science Institutes (ENFSI), *Guideline for Evaluative Reporting in Forensic Science* (2015), https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf; Kaye et al., *The New Wigmore: A treatise on Evidence: Expert Evidence*; D. V. Lindley, "A Problem in Forensic Science," *Biometrika* 64, no. 2 (1977), https://doi.org/10.1093/biomet/64.2.207; Parker, "A Statistical Treatment of Identification Problems."; Robertson, Vignaux, and Berger, *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*; Glenn Shafer, "Lindley's Paradox," *Journal of the American Statistical Association* 77, no. 378 (1982), https://doi.org/10.1080/01621459.1982.10477809.

[239] Including the Netherlands Forensic Institute, the School of Criminal Justice, University of Lausanne, and the Swedish National Forensic Center (see, for example, A. Nordgaard et al., "Scale of Conclusions for the Value of Evidence," *Law, Probability and Risk* 11, no. 1 (2011), https://doi.org/10.1093/lpr/mgr020; R. Marquis et al., "Discussion on How to Implement a Verbal Scale in a Forensic Laboratory: Benefits, Pitfalls and Suggestions to Avoid Misunderstandings," *Science & Justice: Journal of the Forensic Science Society* 56, no. 5 (Sep 2016), https://doi.org/10.1016/j.scijus.2016.05.009; W. Kerkhoff et al., "Cartridge Case and Bullet Comparison: Examples of Evaluative Reporting," *Association of Firearm and Toolmark Examiners Journal* 49, no. 2 (2017), https://nrl.northumbria.ac.uk/id/eprint/35853; A. van Es et al., "Implementation and Assessment of a Likelihood Ratio Approach for the Evaluation of LA-ICP-MS Evidence in Forensic Glass Analysis," *Science & Justice: Journal of the Forensic Science Society* 57, no. 3 (May 2017), https://doi.org/10.1016/j.scijus.2017.03.002.)

[240] See for example C. G. G. Aitken and D. A. Stoney, *The Use of Statistics in Forensic Science* (London: CRC Press, 1991); I. W. Evett and Weir. B. S., *Interpreting DNA Evidence* (Sunderland, MA: Sinauer, 1998); C. Champod, I. Evett, and B. Kuchler, "Earmarks as Evidence: A Critical Review," *Journal of Forensic Sciences* 46, no. 6 (2001), https://doi.org/10.1520/JFS15146J; Silvia Bozza et al., "Probabilistic Evaluation of Handwriting Evidence: Likelihood Ratio for Authorship," *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57, no. 3 (2008), https://doi.org/10.1111/j.1467-9876.2007.00616.x.

[241] Criticism of this approach/paradigm have been stated. For details and discussion, see Shafer, "Lindley's Paradox."

Box 2.3: Evidential strength in a handwriting case (likelihood ratio paradigm)

The law of likelihood implies that for a set of features observed in the evidence (E), if the chance of observing these features if H_1 (Mr. X wrote the questioned document) is true is larger than the chance of observing these features if H_2 (someone else wrote the questioned document) is true, then this evidence supports H_1 over H_2.

Evidential strength, as defined by Royall,[242] is based on probability. To be more specific, it is based on two probabilities, and the task of the FDE is essentially to provide a judgement on these probabilities based on observation E and the possible causes of E, H_1, and H_2.

The judgement can be based on data and personal belief, although the FDE must be explicit in whether the judgement is based on data, personal belief, or both.

For example, if the observations are that "there is a very close correspondence between the question and known documents," the FDE may judge that they expect this if Mr. X wrote the questioned document (H_1), and consequently, that there is a high enough probability to make this observation. In addition, if an FDE thinks that the questioned document handwriting is a relatively rare type in some population of writers, then the FDE does not expect to see this type if someone other than Mr. X wrote the questioned document (H_2). The FDE consequently thinks that there is a small probability of observing this handwriting type in the population of writers that they are considering. The fact that the likelihood under H_1 is judged to be larger than the likelihood under H_2 implies the observations are evidence that H_1 is true relative to H_2. The strength of the evidence depends on the size of the difference between these two likelihoods.

If there is a relevant quantitative database available that can be used to estimate the probabilities as rates (e.g., 99 in 100 and 1 in 100, respectively), the FDE can provide a quantitative judgement of 99 for the evidential strength (i.e., the likelihood under H_1 is 99 times larger than the probability under H_2).

If there are no data (or no relevant data), then the FDE can still assess the evidential strength based on qualitative subjective or personal probabilities. The FDE thinks the probability of H_1 to be quite high and the probability of H_2 to be quite low. Subsequently, the FDE can infer that the observations are much more probable under H_1 than under H_2.

Even if the FDE cannot provide individual probabilities, they may be able to compare them directly and judge, even without knowing the values of the probability itself, that E is much more probable under H_1 than under H_2.

There are several approaches on the proper domain of mathematical probability,[243] of which the frequentist (probability based on the frequency of occurrence of an event) and the subjective or

[242] R. Royall, *Statistical Evidence: A Likelihood Paradigm*, Monographs on Statistics and Applied Probability, (Chapman & Hall/CRC Press LLC, 1997).

[243] A. Hájek, "Interpretations of Probability," E. N. Zalta ed. *The Stanford Encyclopedia of Philosophy*, 2012, https://plato.stanford.edu/archives/win2012/entries/probability-interpret.

Bayesian approaches are the most prominent in the forensic sciences. Among forensic statisticians, there is a continuous, strong, and active discussion about the concept of probability and how to apply it in forensic science. This discussion is fostered by the fundamental differences between the frequentist approach and the Bayesian approach (see box 2.4). This discussion has deep roots in statistical and mathematical science and may never reach a solution that satisfies all those contributing to the discussion. It is important, however, for every person working in forensic science (e.g., forensic scientists) or using forensic science (e.g., judges and juries) to have a basic understanding of what probability is and what types of probability are used in each aspect of forensic testimony and reporting. Essentially, there is a common agreement among statisticians, legal scholars, and scientists—advocating either approach to evidence interpretation—that various types of probabilistic reasoning are the foundation for the science of forensic individualization. Differences between the two approaches should not prompt non-statisticians to dismiss probability as the core concept in forensic science evidence evaluation.

Box 2.4: Bayesian approach and frequentist approach

As noted in the main text, the Bayesian approach and the frequentist approach differ in their definition of probability and the mathematical model they use to model reality. This box describes some of the differences between the approaches in more detail.

- In the Bayesian approach, probability is defined as a degree of belief, which is dependent on the available information, dependent on the person (personal/subjective), and has no "true" value. By contrast, the frequentist approach views probability as a frequency of occurrence (i.e., a relative frequency). It does have a true value (i.e., the population value) and does not depend on the person (objective).
- In the frequentist approach, probability is understood as an event occurring by chance. It is usually applied to sampling experiments on well-defined populations and is used to discuss the rate at which certain features are encountered in the specified population.
- For non-recurring events like "the event that John threatened his brother" or "the event that the suspect is guilty," the Bayesian approach is better equipped than the frequentist approach. The frequentist approach requires that one conduct an experiment because probability is understood to be the frequency of occurrence. For non-recurring events, this poses a challenge. The concept of a hypothetical thought experiment has been developed as a pragmatic solution to this issue (see appendix 2A).
- Generally speaking, Bayesian methods work well for Bayesian probabilities and frequentist methods work well with frequentist probabilities. When combining Bayesian and frequentist methods, one must exercise caution to not end up with an ad hoc methodology that offers none of the advantages of either paradigm.

Given the complexity of using probabilistic reasoning to interpret handwriting evidence, FDEs require a basic knowledge of the differences and uses of the two types of probability and clarity about what is meant by each. Teaching these concepts should include an overview of each paradigm without recommending one over the other, because each serves a different purpose. An

FDE's choice of which particular type of probability to use should reflect the type of statement the FDEs wish to make and the audience to whom they are presenting the evidence (e.g., a judge, jury, or reader of a written report). Research is needed to better understand how to best convey these concepts to FDEs and to consumers of handwriting examinations.

2.3.2.1. Propositions

Regardless of the approach an FDE uses, when evaluating evidence there must be at least two mutually exclusive competing propositions (or hypotheses). It should be noted that although the conventional approach may also use competing propositions, they may not be as explicitly detailed as in other approaches. For instance, FDEs using the conventional approach may default to using an alternative proposition that someone else in the population wrote the text. Mutually exclusive means that there should be no overlap, implying that the propositions being compared cannot both be true at the same time. Ideally, the propositions should reflect the positions that will be presented in court and argued by opposing parties. When this is not possible, however, the FDE may suggest the most reasonable and relevant propositions based on task-relevant contextual information. As discussed in section 2.1, care should be taken that the information necessary to formulate the propositions does not bias the examination.

The propositions explicitly determine the type of information needed, which may differ from case to case. The propositions also define the relevant population with respect to the case under consideration. For example, in the hypothetical case of a suicide note that might have been forged by the twin brother and no one else (section 2.1.7), the two propositions are that the deceased wrote the note (H_1) and that the brother wrote the note (H_2).[244] In this case, H_1 and H_2 define what information is needed to perform the examination. These propositions require reference handwriting from both brothers.

If, on the other hand, the alternative proposition were not confined to the brother but to a person from the community where the suspect lives, the two competing propositions would be that the deceased wrote the note (H_1) and that another person from the community wrote the note (H_2).

The propositions could be refined further. Perhaps W_1 wrote the note trying to disguise his handwriting, or perhaps he wrote it in his natural handwriting. If someone else wrote the note, perhaps that individual was an elementary school classmate of the deceased and thus might share similar writing characteristics.[245]

The ENFSI *Guideline for Evaluative Reporting in Forensic Science*[246] provides recommendations for implementing the subjective likelihood ratio approach. It states that the conclusion of the examination should follow the principles of balance, logic, robustness, and transparency. The conclusion should express the degree of support provided by the forensic findings for one proposition versus the specified alternative(s). The degree of support relates to

[244] An example of propositions that are not mutually exclusive would be that the deceased wrote the note (H_1) and that someone living in the house of the deceased wrote the note (H_2). If H_1 is true, this implies that H_2 is true as well.

[245] For further discussion of formulating propositions for investigation and evaluation, see T. Hicks et al., "The Importance of Distinguishing Information From Evidence/Observations When Formulating Propositions," *Science & Justice: Journal of the Forensic Science Society* 55, no. 6 (Dec 2015), https://doi.org/10.1016/j.scijus.2015.06.008, https://www.ncbi.nlm.nih.gov/pubmed/26654089; Jackson, Aitken, and P., *Case Assessment and Interpretation of Expert Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*.

[246] European Network of Forensic Science Institutes (ENFSI), *Guideline for Evaluative Reporting in Forensic Science*.

the magnitude of the likelihood ratio. A likelihood ratio may be expressed by a number or a verbal equivalent according to a specified scale of conclusions.[247] The guideline also discusses propositions,[248] with several important aspects to be considered, including the hierarchy of propositions (sub-source/source/activity/crime) and the importance of an alternative proposition. The alternative proposition is usually that some other writer is the source of the writing sample. This proposition is not formal or explicit in a strict statistical sense, in part because no reference is made to the relevant population. In practice, defining and assessing the relevant population is difficult; however, for the sake of transparency the population being drawn from should be disclosed to include past experience with this population. Although the level in the hierarchy of propositions is not as obvious for handwriting as for some other types of evidence, it should be made explicit when an FDE moves beyond source-level propositions toward the activity-level propositions.[249]

Recommendation 2.5: A forensic handwriting examination should be based on at least two mutually exclusive propositions relevant to the examination(s) requested. These propositions should be explicitly taken into account in the interpretation of the handwriting evidence and included in the conclusion, report, and testimony.

2.4. Research Needs

The Working Group has identified several research areas that could improve the application and accuracy of forensic handwriting examination. First, more research is needed to identify and validate FDE claims about the opinions they can render in handwriting examination (see section 2.2). Examples of such claims, given a sufficient quantity and quality of questioned and comparison material, include that FDEs can

- Provide an opinion on whether the writer of the comparison material wrote the questioned material when both materials are uppercase print;
- Provide an opinion on whether the writer of the comparison material wrote the questioned material when both materials are lowercase cursive;
- Provide an opinion when the comparison material and or the questioned material are non-originals; and
- Provide an opinion on whether the questioned and comparison materials are the products of simulation or disguise behavior.

[247] European Network of Forensic Science Institutes (ENFSI), *Guideline for Evaluative Reporting in Forensic Science*, 16.

[248] European Network of Forensic Science Institutes (ENFSI), *Guideline for Evaluative Reporting in Forensic Science*, 11–15.

[249] R. Cook et al., "A Hierarchy of Propositions: Deciding Which Level to Address in Casework," *Science & Justice* 38, no. 4 (1998), https://doi.org/10.1016/S1355-0306(98)72117-3; I. W. Evett, G. Jackson, and J. A. Lambert, "More on the Hierarchy of Propositions: Exploring the Distinction Between Explanations and Propositions," *Science & Justice* 40, no. 1 (2000), https://doi.org/10.1016/s1355-0306(00)71926-5.

Although studies have been conducted and reported,[250] the full comprehensive list of claims is unknown, making it difficult to assess whether there is empirically valid evidence to support their use. Examination methods should be based on empirically supported data.

Recommendation 2.6: The forensic document examiner community should consider the claims made by forensic document examiners and then conduct empirical studies in collaboration with the research community to characterize the extent of scientific support for those claims.

Second, as noted in section 2.3, FDEs could benefit from sample data from different locales and population groups. The term population can represent either the general population or a more specific population of interest or relevance (subgroup). Well-constructed databases containing a large amount of writing, where all the features of interest have been measured, can provide insight into and estimates of the frequencies and interdependences of salient features in the studied populations (i.e., the frequency of occurrence of inter-writer and intra-writer features and combinations of features). Frequency estimates from such data could provide a more objective foundation for FDE assessment of the features and their relative value compared with personal experience–based judgements.

One currently available database consists of 1,500 handwriting and hand printing samples obtained from the general public with estimates of the frequency of occurrence of features.[251] Although having representative data for the population of interest in a given case is ideal, even if a given database is not a random sample from the relevant population, it may still have some value for the examination. That is, although an explicit database is always preferred over the implicit database in the mind of the FDE, some information may be better than no information. The relevance and use of any given database should be determined by the FDE on a case-by-case basis, and there should be transparency in this decision-making process.

Research about baseline occurrences of particular features in a population should include studies addressing

- **Occurrence of features by geographic area**. Such studies should address regional commonalities in writing attributes (i.e., class characteristics).
- **Occurrence of combinations of features**. Studies of feature combinations should address both commonly occurring and rarely occurring combinations of letters, numbers, or other distinguishing characteristics of writing.
- **Identification of rarely occurring features**. Rarely occurring features such as character forms, diacritics, or other sources of variation should be addressed.

[250] See for example C. Bird, B. Found, and D. Rogers, "Forensic Document Examiners' Skill in Distinguishing Between Natural and Disguised Handwriting Behaviors," *Journal of Forensic Sciences* 55, no. 5 (Sep 2010), https://doi.org/10.1111/j.1556-4029.2010.01456.x; B. Found, J. Sita, and D. Rogers, "The Development of a Program for Characterising Forensic Handwriting Examiners' Expertise: Signature Examination Pilot Study," *Journal of Forensic Document Examination* 12 (1999); Moshe Kam et al., "Signature Authentication by Forensic Document Examiners," *Journal of Forensic Sciences* 46, no. 4 (2001), https://doi.org/10.1520/jfs15062j; Jodi Sita, Bryan Found, and Douglas K. Rogers, "Forensic Handwriting Examiners' Expertise for Signature Comparison," *Journal of Forensic Sciences* 47, no. 5 (2002), https://doi.org/10.1520/jfs15521j.

[251] Johnson et al., "Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics."

- **Identification of characteristics common among and specific to population subgroups**. These studies should include characteristics that may identify writers as members of foreign populations, non-native writers, or writers who are not literate in specific writing systems.

Recommendation 2.7: The forensic document examiner community, in collaboration with researchers, should design and construct publicly available, large databases of representative handwriting features to facilitate research in and improve the accuracy of handwriting examination.

Finally, the Working Group identified several additional key priorities for feature interpretation research studies:

- Writing complexity. These studies should define how complexity is measured and the level to which complexity is sufficient for meaningful comparisons for all types of writing, like hand printing, numerals, signatures, or foreign writing systems.
- Developing methods of quantifying and measuring inter-writer and intra-writer variability. Such studies should include cross-cultural writing and longitudinal studies of changes in writing across time and studies of writing characteristics that arise in the absence of formal instruction in cursive writing and penmanship.
- Amount of writing required to reach a conclusion about the writership of the questioned writings. Studies should include the degree of writing complexity required to establish the presence or absence of diagnostic features, the minimum quantity of writing needed to form reliable opinions, cross-cultural studies, and studies specifically addressing writing forms like numerals, signatures, initials, and hand-printed materials.
- Comparability of types of writing. These studies should include forms of writing like initials, signatures, hand printing, and foreign writing.
- Relevant information (features) identified in writing samples and the extent of the consistencies in how such information is interpreted. These studies should address the extent to which information in the written materials has the potential to reliably indicate whether the writing is genuine or non-genuine (i.e., disguised, traced, or produced by some other method of simulation) and how consistently such information is used to establish the writership of a questioned writing.

These studies should be performed where participants have access to the standard tools and equipment commonly used by members of the field to investigate whether findings obtained in an experimental laboratory are replicated in a document examination laboratory setting.

2.5. Automated Systems

This section describes automated pattern-matching methods based on statistics and computer science that might supplement FDE evaluations. Approaches to automated handwriting

identification and verification[252] have been studied and developed since the mid- to late 1980s.[253] Franke and colleagues[254] took a leading role during this early stage and based much of their development on semi-automated systems, like Forensic Information System for Handwriting (FISH)[255] and later a version of FISH called WANDA.[256] These early systems were parallel efforts to develop offline handwriting recognition systems.[257]

Pattern recognition is an important example of this early work; however, the group[258] did not base their efforts on conventional handwriting features used by FDEs. Instead, they developed new sets of features based on computer vision and vector quantization. Building on these early proof-of-concept approaches, the NIJ funded a series of research projects, led by Sargur Srihari at the Center of Excellence for Document Analysis and Recognition (CEDAR), to develop an automated system based on features derived from those used by FDEs to study the foundations of questioned document analysis.[259]

Automated handwriting feature recognition systems remain the purview of large public laboratories or engineering departments within universities. A 2014 survey[260] of 95 FDEs asked, "If you use an automated handwriting system, which one (or more) do you use?" Seventy-three percent responded that they had not used any of the available systems. Of the systems reported to have been used by the survey participants, CEDAR-FOX (or the interactive version, CEDAR-iFOX), Forensic Language-Independent Analysis System for Handwriting Identification (FLASH ID),[261] and FISH were the most common.

Automated handwriting feature recognition systems have been deployed to support the basic tenets of handwriting, to facilitate FDE decision making with regard to feature selection, and to

[252] In the field of handwriting biometrics where automated systems are used to analyze and compare handwriting, the term "writer identification" is used when establishing the identity of an individual from a given list (a 1:*N* comparison), and "writer verification" is used when a 1:1 comparison is undertaken to verify the identity of a specific writer. L. Schomaker, "Writer Identification and Verification," in *Advances in Biometrics*, ed. N. K. Ratha and V. Govindaraju (London: Springer, 2008), 248.

[253] Réjean Plamondon and Guy Lorette, "Automatic Signature Verification and Writer Identification—the State of the Art," *Pattern Recognition* 22, no. 2 (1989), https://doi.org/10.1016/0031-3203(89)90059-9.

[254] K. Franke et al., "FISH-New: A Common Ground for Computer-Based Forensic Writer Identification," *Forensic Science International* 136, no. S1-S432 (2003).

[255] H. W. Eiserman and M. R. Hecker, "FISH-computers in handwriting examinations" (44th Annual Meeting of the American Society of Questioned Document Examiners, Savannah, GA, 1986).

[256] K.Y. Franke et al., "WANDA: A Common Ground for Forensic Handwriting Examination and Writer Identification," *ENFHEX News—Bulletin of the European Network of Forensic Handwriting Experts* 1, no. 4 (2001), https://www.academia.edu/26020856/WANDA_A_common_ground_for_forensic_handwriting_examination_and_writer_identification.

[257] H. E. S. Said, T. N. Tan, and K. D. Baker, "Personal Identification Based on Handwriting," *Pattern Recognition* 33, no. 1 (2000), https://doi.org/10.1016/s0031-3203(99)00006-0.

[258] Franke et al., "WANDA: A Common Ground for Forensic Handwriting Examination and Writer Identification."; Franke et al., "FISH-New: A Common Ground for Computer-Based Forensic Writer Identification."; Said, Tan, and Baker, "Personal Identification Based on Handwriting."

[259] S.N. Srihari, *Computational Methods for Handwritten Questioned Document Examination* (2010), https://www.ncjrs.gov/pdffiles1/nij/grants/232745.pdf.

[260] J. P. Jones, "The Future State of Handwriting Examinations: A Roadmap to Integrate the Latest Measurement Science and Statistics" (AAFS Annual Meeting, Seattle, WA, February 20, 2014 2014).

[261] Saunders, Davis, and Buscaglia, "Using Automated Comparisons to Quantify Handwriting Individuality."; D. T. Gantz and M. A. Walch, "FLASH ID Handwriting Derived Biometric Analysis Software" (NIST Measurement Science and Standards in Forensic Handwriting Analysis Conference, 2013).

study error rates compared with human FDEs.[262] These efforts underscore the potential of these systems to validate writership claims.

2.5.1. The Early Years of Automated Systems

Early efforts focused on estimating the chance (i.e., the frequentist probability) of observing two writers in a given population with non-unique writing profiles. If this chance were zero, then the reasoning followed that every individual in the population would have a unique writing profile. The first of these projects attempted to statistically demonstrate that each writer possessed a unique writing profile in the general U.S. population of writers.[263]

There was also a focus on developing strategies to perform a large number of comparisons between handwriting exemplars. Srihari and colleagues[264] conducted a study to test the principle of individuality. The researchers built an automated writer identification system to use as a comparison method for examining writing samples in the context discussed in chapter 1. Samples from 1,500 individuals from the general U.S. population, including men and women of different ages and ethnicities, were collected and entered into a database. Each individual provided three handwritten samples that captured the various attributes of the written English language, including document structure (e.g., word and line spacing, line skew, margins), positional variations of the letters (e.g., each letter in the initial, middle, and terminal positions of a word), and letter and number combinations (e.g., ff, tt, oo, 00). The software program CEDAR-FOX was developed to extract macro-features (e.g., slant, word proportion, measures of pen pressure, writing movement, and stroke formation) from the entire document based on one paragraph and one word in the document. It also extracted micro-features (e.g., gradient, structural, and concavity features) at the character level of the document.

Applying CEDAR-FOX to handwriting from twins and non-twins, Srihari et al.[265] found that handwriting of twins is harder to distinguish than that of non-twins and that the handwriting of identical twins is harder to distinguish than that of fraternal twins. The system determined, based on a half-page of extended handwriting,[266] that the writer identification error was 13% for twins compared with 4% for non-twin samples. Srihari et al. concluded that with further improvements, machine-based handwriting verification systems can achieve accuracy levels comparable to expert FDEs.

Although numerous studies have examined handwriting identification and verification systems, Srihari et al.'s study was the first attempt at relating the results of the identification system to the concepts of uniqueness and individuality in handwriting. [267] Koehler and Saks[268] noted a concern that demonstrating uniqueness would require, among other things, a census of all writing profiles. The best a statistician can do without looking at every individual in a given population,

---

[262] Srihari, Huang, and Srinivasan, "On the Discriminability of the Handwriting of Twins."

[263] Srihari et al., "Individuality of Handwriting."

[264] Srihari et al., "Individuality of Handwriting."

[265] Srihari, Huang, and Srinivasan, "On the Discriminability of the Handwriting of Twins."

[266] Twins' handwriting samples were collected by the U.S. Secret Service using the same text as in the CEDAR letter. They are available for download from https://www.cedar.buffalo.edu/~srihari/papers/JFS2008-color.pdf (Srihari, Huang, and Srinivasan, "On the Discriminability of the Handwriting of Twins.")

[267] Srihari et al., "Individuality of Handwriting."

[268] Koehler and Saks, *Individualization Claims in Forensic Science: Still Unwarranted*.

is to estimate the chance of observing two indistinguishable individuals (with respect to a given comparison methodology) who are randomly selected from the population. This issue is not unique to handwriting.[269]

2.5.2. Automated Systems to Support Handwriting Examinations

During early efforts, the FISH and CEDAR-FOX systems demonstrated that it is possible to use a computer-assisted system in forensic identification of source problems associated with questioned document analysis.[270] Although the success of these methods in providing evidence for the tenet that every individual possesses a unique handwriting profile is debatable, these systems demonstrated that it is possible to identify the writer of a questioned document (in a biometric sense) with high accuracy.[271] Toward the end of this stage of development, the focus shifted to "how to present and interpret" the results of these systems to a decision maker.[272] These types of questions tend to rely on a likelihood ratio approach, as typified by the researchers and experts associated with the British Forensic Science Service and the Netherlands Forensic Institute and the forensic science experts in evidence interpretation at the University of Lausanne and government FDEs in Australia.[273]

The first semi-automated approaches for handwriting evidence quantification appear to have been developed by Bozza et al.[274] This formal Bayesian approach focused on summarizing the evidence to support a decision maker in deciding between two forensic propositions: "The suspect wrote the questioned document versus someone else wrote the questioned document."

The method developed a likelihood ratio for writership of a questioned document based on closed loop "o"s. Although the method has been extended to other types of letters in later papers,[275] to the Working Group's best knowledge, this is the only statistically rigorous and formal evidence interpretation approach for handwriting analysis.

In machine learning, the logic of the computer program is determined from examples rather than defined by the programmer. Earlier machine learning approaches required the programmer to design algorithms to compute features/characteristics. In a new development called deep learning, the system itself learns the internal representation. Deep learning has proved useful for performing discrimination in tasks like speech recognition, computer vision, natural language processing, and recommendation systems.[276]

---

[269] M.J. Saks and J. J. Koehler, "The Individualization Fallacy in Forensic Science Evidence," *Vanderbilt Law Review* 61, no. 1 (2008).

[270] See C. P. Saunders et al., "Construction and Evaluation of Classifiers for Forensic Document Analysis," *Annals of Applied Statistics* 5, no. 1 (2011); M. L. Bulacu, "Statistical Pattern Recognition for Automatic Writer Identification and Verification" (Ph.D. University of Groningen, 2007).

[271] Srihari, Huang, and Srinivasan, "On the Discriminability of the Handwriting of Twins."

[272] J. J. Miller et al., "A Set of Handwriting Features for Use in Automated Writer Identification," *Journal of Forensic Sciences* 62, no. 3 (May 2017), https://doi.org/10.1111/1556-4029.13345, https://www.ncbi.nlm.nih.gov/pubmed/28054339.

[273] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83.

[274] Bozza et al., "Probabilistic Evaluation of Handwriting Evidence: Likelihood Ratio for Authorship."

[275] R. Marquis et al., "Handwriting Evidence Evaluation Based on the Shape of Characters: Application of Multivariate Likelihood Ratios," *Journal of Forensic Sciences* 56 Suppl 1 (Jan 2011), https://doi.org/10.1111/j.1556-4029.2010.01602.x.

[276] L. Deng, G. Hinton, and B. Kingsbury, "New Types of Deep Neural Network Learning for Speech Recognition and Related Applications: An Overview" (2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, 2013); Alexandros Karatzoglou and Balázs Hidasi, "Deep Learning for Recommender Systems" (Proceedings of the Eleventh ACM Conference on Recommender Systems, Como, Italy, Association for Computing Machinery, August 2017 2017).

Bozza's approach showed that it was possible to characterize uncertainty of an FDE's conclusion in the form of an ad hoc, machine learning–based likelihood ratio.[277] The automated approaches to handwriting identification show that it is possible to use likelihood-based methods for writer identification and verification tasks. However, the performance (in terms of computational complexity and accuracy) of the automated approaches to closed set identification must significantly improve to be useful in forensic document examination. It remains unclear how best to measure performance in automated forensic identification of source problems. Nonetheless, automated systems have great potential for improving performance in terms of the computational speed of the algorithms and accuracy; new developments in this field should be incorporated into the examination process as they become available.[278]

Automated systems can reduce subjectivity associated with certain human factors such as sufficiency determination, quality decisions, feature selection and extraction, feature matching, and interpretation. However, it is important to recognize that automated systems can present the FDE with other challenges. For example, with the exception of automated signature verification competitions sponsored by the International Conference on Document Analysis and Recognition (ICDAR) (2011–2013), studies[279] have used different sets of known signature or handwriting exemplars to serve as known cases. The absence of a standard set of known signature or handwriting exemplars makes it difficult to compare the value of different automated systems. In addition, most automated feature identification systems are designed to perform well with respect to their intended purpose. Most systems are geared for investigative work to facilitate large-scale processing of questioned documents; they focus on closed set identification of sources. However, the systems have not been tested to determine if they can correctly answer specific questions about writership in actual casework where issues of simulation and disguise are regularly encountered.

The majority of published studies of automated handwriting identification systems are based on comparisons of documents with similar content. Typical examples of content are the "London Letter," "Dear Sam," or repetitions of common phrases.[280] These whole sets of writing samples are then compared using an automated system designed to address the task of interest, typically writer verification or writer identification.[281] One early concern, pointed out by Bulacu et al.,[282] is that ideal features used in an automated system should not depend on the underlying content.

A common automated approach for analyzing handwriting evidence is to develop algorithms for computing features of handwritten characters and algorithms to determine layout characteristics

---

[277] See Saunders et al., "Construction and Evaluation of Classifiers for Forensic Document Analysis."

[278] "Transdisciplinary Research in Principles of Data Science (TRIPODS)," updated March 19, 2019, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505347.

[279] Said, Tan, and Baker, "Personal Identification Based on Handwriting."; Srihari et al., "Individuality of Handwriting."; Srihari, Huang, and Srinivasan, "On the Discriminability of the Handwriting of Twins."

[280] S.N. Srihari et al., *Individuality of Handwriting* (2001), 7, https://www.ncjrs.gov/pdffiles1/nij/grants/190133.pdf; Somaya Al-Maadeed, "Text-Dependent Writer Identification for Arabic Handwriting," *Journal of Electrical and Computer Engineering* 2012 (2012): 4, https://dx.doi.org/10.1155/2012/794106.

[281] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer Identification Using Edge-Based Directional Features," *Proceedings of the Seventh International Conference on Document Analysis and Recognition* 2 (2003): 937, https://www.ai.rug.nl/~mbulacu/icdar2003-bulacu-schomaker-vuurpijl.pdf.. Writer verification is a task focused on doing one-to-one comparisons between handwriting samples with the goal of minimizing the false association and false exclusion rates. "Identification" is the term used in pattern recognition, but it should be more properly thought of as writer recommendation.

[282] Bulacu, "Statistical Pattern Recognition for Automatic Writer Identification and Verification."

(e.g., spacing between words and lines). The automated system first generates a similarity metric between known and questioned handwriting using the computed characteristics. Using probability distributions of the score—as determined from handwriting samples collected from a population assumed to be representative of the United States—the system computes a score-based likelihood ratio. It is also possible to determine the system error rate by determining whether the likelihood ratio is above/below 1 when the questioned and known writings are from same or different individuals, respectively. The scores produced showed over 95% accuracy,[283] which provided support for admitting handwriting testimony in *Daubert*[284] and *Frye*[285] hearings.[286]

One particular study involving handwriting (not signatures) showed that FDEs performed better than certain types of automated systems.[287] Most automated systems for forensic handwriting analysis are designed for different tasks, either to construct different types of values of the evidence or to serve as recommender systems to suggest in what order FDEs should compare knowns from different writers to a given source. However, in the context of biometrics and signature verification, at least one study of signatures directly compared an automated signature verification system with FDEs showing automated signature verification systems to perform similarly to human FDEs.[288]

As with human experts, the error rate in computer models depends on the difficulty of the task and reliable estimates of source variability. Depending on the task and the specifics of the automated systems, writer identification systems perform as well as human experts in certain metrics.[289] In the absence of empirical research, it is unclear whether automated systems return inconclusive decisions at the same rate as expert FDEs. Such a comparison is made difficult, if not impossible, given that it is rare to design a system that returns inconclusive results. Unlike expert handwriting or signature identification, automated systems are not subject to motivational or confirmation biases, nor task-irrelevant contextual information, which might inflate error rates.

[283] Srihari et al. defined identification accuracy as "measured against the number of writers considered in three separate sets of experiments using macro-features, micro-features, and their combinations." (Srihari et al., "Individuality of Handwriting.")

[284] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.

[285] *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

[286] *United States v. Prime*, 220 F. Supp. 2d 1203 (W.D. Wash. 2002); *Pettus v. United States*, 37 A.3d 213 (D.C. 2012).

[287] Srihari, Huang, and Srinivasan, "On the Discriminability of the Handwriting of Twins."

[288] M. I. Malik et al., "Man vs. Machine: A Comparative Analysis for Signature Verification," *Journal of Forensic Document Examination* 24 (2014), https://doi.org/10.31974/jfde24-21-35.

[289] Malik et al., "Man vs. Machine: A Comparative Analysis for Signature Verification."

Prior research (cited above) on error rates associated with automated handwriting and signature recognition systems focused on different pattern recognition tasks. Most concentrated on common but unknown sources or closed set identification (i.e., limited reference population). In general, error rates were functions of the document sizes (i.e., volume of writing), the number of samples in the candidate list (i.e., returned from a search), or number of enrolled writers in the database.[290]

2.5.3. The Future of Automated Systems

As expertise in questioned document analysis becomes rarer, automated systems can provide a critical system of tools for writership analysis. Several systems provide capabilities for comparing handwriting samples, including FLASH ID and CEDAR-FOX. These systems provide a list of possible writers of a questioned document. Other systems, like WANDA and FISH, also provide markup and process documentation for questioned document analysis. Hands-on use of the tools requires one-on-one interaction between the trainer and trainee. Furthermore, the software may be improved by using case-specific training samples provided by the FDE. More research is needed to interpret the results of the system (e.g., in terms of a likelihood ratio).

In a deep learning approach to forensic document examination, handwriting characteristics used to compare questioned and known documents are determined by the system itself, rather than by an FDE or the programmer. In performing a handwriting examination, features are the input, and the deep learning methods provide very flexible models for learning the classification rules for feature analysis. The computational requirements for machine learning algorithms for complex evidence forms, such as handwritten documents, are high. Typically, there are billions of parameters that need to be learned (or optimized) from the limited number of control/training samples. It is expected that the major advances in cloud computing (Amazon provides fast processors useful for deep learning, called graphics processing units) and software systems (Google released Tensorflow[291] into the public domain) will make it possible to develop such tools in the near future (3 to 5 years). This approach will be inherently interdisciplinary, requiring collaborations between the broadly defined data science community and FDEs, especially in the design, testing, and evaluation phases of the research.[292] As automated systems for feature assessment and interpretation grow in number and reliability, FDEs should be open to including them as components of their examination of casework.

[290] National Science Foundation (NSF), "Transdisciplinary Research in Principles of Data Science (TRIPODS)."; Marcus Liwicki et al., "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)" (2011 International Conference on Document Analysis and Recognition, 2011); Muhammad Imran Malik and Marcus Liwicki, "From Terminology to Evaluation: Performance Assessment of Automatic Signature Verification Systems" (2012 International Conference on Frontiers in Handwriting Recognition, Bari, Italy, IEEE, 2012); Muhammad Imran Malik et al., "ICDAR 2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp 2013)" (12th International Conference on Document Analysis and Recognition, 2013).

[291] An open-source software library for numerical computation. See https://www.tensorflow.org/ ("Tensor Flow home page," 2019, accessed May 6, 2020, https://www.tensorflow.org/.).

[292] M. Liwicki, M. I. Malik, and C. E. H. Berger, "Towards a Shared Conceptualization for Automatic Signature Verification," in *Advances in Digital Handwritten Signature Processing*, ed. D. Impedovo G. Pirlo, and M. Fairhurst (Singapore: World Scientif, 2014).

Recommendation 2.8: The forensic document examiner community should collaborate with the computer science and engineering communities to develop and validate applicable, user-friendly automated systems.

Appendix 2A: Probability and Statistical Reasoning

This appendix introduces some basic ideas of probability and statistical reasoning. The meaning of probability is explained, and probabilities are described for propositions like H_1 and H_2. Then, the appendix explains how these can be used to assist the fact finder.

Probability

In mathematics, probabilities are numbers that obey a few axioms.[293] One standard axiom requires probabilities to be single numbers between zero and one. A probability of zero for a proposition means that it is not true. A probability of 1 means that the proposition is true. Probability is often expressed as a percentage or as a natural frequency. Probabilities of 0.75, 75%, or 75 out of 100 are all equivalent expressions. Probability can also be presented in terms of odds. If the probability is 75%, the odds are expressed as 75:25 (or, equivalently, 3:1).[294]

The mathematics of probability has its roots in studies of games of chance. Today, the mathematical structure for the probabilities of events, such as the outcomes for card games, lotteries, radioactive decay, inheritance of genes, and measurements of chemical and physical properties is well understood. To apply probability to forensics, one must determine whether the same calculus applies to things other than the outcomes of inherently stochastic or random processes. Can it be used to quantify the degree of certainty or belief that an expert (or a judge or jury) might express in the truth of statements like "Person X was the source of trace evidence"?

The frequentist school defines probability as the so-called long-term relative frequency of an event. This definition implies a repeated measurement of the event by means of an experiment or other form of data collection. As an example, consider the statement "there is a low probability that a certain writer writes the number "8" in a particular way." This can be understood as a statement about the occurrence of this 8 in a population of writings made up of that specific individual's writings. A low probability implies that only a small amount of the writing samples (e.g., 1 out of 100) would contain an 8 that is similar in a particular way to the observed 8 in question. A limitation of the frequency-based school, in its most basic and strict form, is that it does not easily permit probabilities to be assigned to non-recurring events.[295]

---

[293] A. N. Kolmogorov, *Foundations of the Theory of Probability* (New York: Chelsea Publishing Company, 1933).

[294] Various studies suggest that most people are better at understanding "natural frequencies" (e.g., 75 out of 100) than probabilities (U. Hoffrage and G. Gigerenzer, "Using Natural Frequencies to Improve Diagnostic Inferences," *Academic Medicine* 73, no. 5 (1998), https://doi.org/11858/00-001M-0000-0025-A092-2.).

[295] However, in most modern applications of this type of probability, the statistician or scientist relies on a concept of a hypothetical random experiment. These hypothetical thought experiments involving an "imaginary long run" (Denny Borsboom, Gideon J. Mellenbergh, and Jaap van Heerden, "Functional Thought Experiments," *Synthese* 130, no. 3 (2002), https://doi.org/10.1023/a:1014840616403.) allow for the application of frequentist statistical techniques to settings involving nonrecurring events. Perhaps one can say confidently that individual W_1 will produce handwriting with certain features a certain fraction of the time and interpret that fraction as a probability that W_1 would have produced a sample with such features on a particular occasion. The variations in the features can be described by a probability function or distribution. But the variability that gives rise to these probabilities pertains to the features—not to proposition H_1 of writership. Either W_1 wrote the questioned specimen, or W_1 did not. One can speak of the probability of the data, or evidence E—the set of features—if W_1 wrote them or if someone else did, but there is no frequency-based interpretation of the proposition H_1 that W_1 was the writer. Expressed in symbols, P stands for the long-run relative frequency of observing a new realization of the evidence (E) in a neighborhood of the observed evidence (e) under a hypothetical sampling experiment implied by H. In short-hand notation, this is typically written as P(H|e). The vertical bar is read as "given" or "conditional on." The "probable" truth of H in light of the realized evidence (e), typically denoted as P(H|e), is not truly a probability in the sense of frequentist probability. To avoid this confusion in statistical discussions, direct or empirical/frequentist probabilities are represented by Latin characters and correspond to either the inherent random nature of a process or a hypothetical experiment-sampling. A similar notion has been invoked to defend reasoning involving subjective probabilities in law (David Kaye, "The Laws of Probability and the Law of the Land," *The University of Chicago Law Review* 47, no. 1 (1979), https://doi.org/10.2307/1599414.).

In contrast, the subjective school of thought does allow for probabilities of non-recurring events. The subjective school of thought conceives of probability as measuring the belief that an individual has in the truth of a proposition, or the occurrence of an event. In this subjective or personal conception, probability is a graded belief for one individual. (e.g., "I am moderately [70 to 80%] confident that the same person wrote both samples"). It is important to understand that this type of probability (i.e., belief) is fundamentally different from the frequentist concept of probability.

The subjective interpretation of probability extends the definition of probability to all propositions about the true state of affairs, where it is used to discuss beliefs concerning the validity of such propositions in a formal or logical manner. The use of personal probabilities in the interpretation and presentation of forensic evidence is typically equated to being logical and coherent in updating one's personal beliefs in light of the empirical evidence. However, one can question the basis for regarding the subjective numbers as mathematical probabilities like the ones defined by the frequentist school of thought. For example, why must an FDE who regards 0.75 as his personal level of partial belief in the proposition that W_1 wrote the document in question also have 0.25 for the partial belief that someone else was the writer?[296]

This exposition is not intended to imply that one definition of probability is correct and that another is wrong. Their range of application simply differs. The subjective conception of probability allows FDEs to have a precise and transparent way of expressing their beliefs, whereas the frequentist conception applies to the rates at which features or objects are observed as a result of a statistical experiment or in a given population. Whatever probability method is employed to interpret and present handwriting evidence, the FDE must be clear about what the probabilities pertain to and measure. It is common to use frequentist probability to discuss the rates at which features or combinations of features occur in a population. It is also common to use subjective probability to characterize beliefs about the rarity of these features in these populations and the inferences that should be drawn from their presence. It is important to keep these two types of probabilities distinct. A forensic scientist may use both types of probability, but a subjective probability not based on comprehensive data from a relevant population should not be presented as if it were a data-driven, frequency-based probability.

Likelihood Ratios, Prior Probabilities, and Source Probabilities

The question of whether observations on a given set of evidence support one hypothesized probability distribution over another is central to statistical inference. The answer to this question is found in the law of likelihood. As Royall[297] describes this relationship, probabilities measure uncertainty whereas likelihood ratios measure evidence. For example, in the simple case of two brothers who are the only conceivable writers of a suicide note, the expert comparing known samples from each brother with the questioned suicide note should have some sense of the relative probability of the evidence in support of one proposition versus an alternative

[296] One argument for demanding that the probabilities an individual would give for every possible proposition should follow the rules for mathematical probabilities is that if personal or logical probabilities are not "coherent" (a technical term meaning that the numbers a person provides for subjective probabilities obey the usual axioms and thus all the rules of probability), then the individual ascribes different probabilities to some logically equivalent propositions. Although students of the foundations of probability and statistics disagree about the force of this argument, especially as applied to individuals with limited time and computational capacities, an expert witness who offered manifestly conflicting assessments of the "probabilities" of conclusions would have little credibility.

[297] Royall, *Statistical Evidence: A Likelihood Paradigm*.

proposition. The writing in the known samples from the surviving twin (W_1) may seem closer to the writing in the suicide note than the writing in the known samples from the deceased twin (W_2). Phrased in statistical terms (see box 2A.1), the observed evidence, typically denoted as *e*, is more probable under one proposition than another: $P(e|H_1) > P(e|H_2)$ corresponds to the observed evidence providing greater support for the proposition that *e* arose under the models in H_1 rather than the models in H_2. If one calls these two probability functions evaluated at the observed evidence (*e*) likelihoods, then the evidence supports H_1 more than H_2 as long as the likelihood ratio $LR = P(E|H_1) / P(E|H_2)$ is greater than 1. If $LR = 1$, the evidence does not let us distinguish between H_1 and H_2. If *LR* is less than 1, the evidence supports H_2 over H_1; the greater the value of *LR*, the greater the support for H_1. In short, the likelihood ratio is a measure of the strength of the evidence. The notion that increasing likelihood $P(e|H_k)$ corresponds to increasing evidentiary support for H_k leads to a school of statistical inference known as the likelihood approach.

Box 2A.1: Terms (and their definitions) used in the statistical expression of likelihood within a formal Bayesian paradigm when evaluating support for one proposition over another

E:	Evidence
e:	Observed evidence
H:	Hypothesis
H_k:	k^{th} hypothesis for how the evidence has arisen
$P(e)$:	The probability of observing the evidence; depending on the context, probability can either be a base frequency of the features or a personal belief
$P(H_k)$:	Prior personal belief, the probability that the conditions of H_k are true
$P(e\|H_k)$:	The probability of *e* occurring given the conditions under H_k is true; depending on the context probability can either be a base frequency of the features or a personal belief given the conditions under H_k is true
$P(H_k\|e)$:	Posterior personal belief, the updated belief of H_k given that *e* has occurred
LR:	Likelihood Ratio
BF:	Bayes Factor

To compute the absolute value of the *LR* for the observed evidence, *e*, the numerical values of $P(e|H_1)$ and $P(e|H_2)$ must be known. Therefore, the *LR* implicitly carries with it a great degree of precision, in the sense that the value of the *LR* (evaluated by a different person who also agrees with the models used in H_1 and H_2) will not be different for the same evidence. This is a very important and appealing aspect of the *LR*, in that when different experts evaluate the same evidence, the value of the *LR* will be fixed. Now, if any uncertainty exists that prevents the exact evaluation of the *LR*, which will be the case in practice, the *LR* ceases to be uniquely defined.

Several different strategies handle this uncertainty, including methods from the formal Bayesian paradigm. Any method that is not the formal Bayesian method of accounting for the uncertainty in the likelihood described below is ad hoc. The resulting statistics from these methods are not defined to be either a formal Bayes factor (*BF*) or an *LR* but an ad hoc solution in between these two well-defined statistics.

However, in forensic statistics, most of the arguments for using likelihoods (whether they are qualitative or quantitative) to evaluate the strength of the evidence come from the formal Bayesian perspective. This framework treats the *LR* (when it is uniquely defined) as measuring the change in belief that the evidence rationally warrants. Again, the observed features in the questioned sample and exemplars are data. The data can make each proposition more or less reasonable than it was before the data were incorporated. The probability $P(H_k)$ before obtaining particular data *E* is known as the prior belief (section 2.2.5 uses the related phrase "base rate"). The belief $P(H_k|e)$ after considering the data is known as the posterior belief.

The precise relationship between the prior and posterior belief is given by a formula known as Bayes' rule. The rule tells us how to update the prior belief in light of the data. When there are only two possible propositions to consider—such as the propositions about the brothers—the increase or decrease in the belief depends on the likelihood ratio. The *LR* is a special case of the general concept of a *BF*, and Bayes' rule dictates that the posterior odds are the prior odds multiplied by the *BF*. A large value of *BF* means that the evidence is powerful—it raises the odds by a large factor.[298] In the Bayesian framework, the *BF* measures the strength of the evidence (just as the *LR* does when there is no uncertainty concerning the nature of how the evidence was generated under the two competing forensic propositions of interest). However, the *BF* may include prior beliefs necessary to characterize how the evidence has arisen under each of the two propositions.[299]

Although FDEs may not be able to provide a quantitative judgement on the likelihood of observing the evidence if the suspect is the writer of the questioned document, they may be able to state that this likelihood is much larger than if a random person, in some population of writers, wrote the questioned document. At a minimum, some qualitative comparisons of the relative support of the data for H_1 over H_2 should be possible. Therefore, the value of the *LR* for these data cannot be obtained, but qualitative likelihoods can be used to obtain a qualitative *BF*. When a qualitative *BF* is used, it carries with it a sense of uncertainty masked by avoiding the specification of prior beliefs used to obtain the *BF* described in the previous paragraph. A qualitative *BF* is a less formal method of expressing the strength of a finding. Therefore, when using a qualitative *BF*, its use should be made explicit to avoid providing a misleading sense of formal rigor to the recipient of this information. The first example in box 2A.2 illustrates the use

[298] Many writers refer to the logarithm of the Bayes factor as the "weight of evidence." (I. .J. Good, *Probability and the Weighing of Evidence* (London: Charles Griffin and Company, 1950); I. .J. Good, "Weight of Evidence and the Bayesian Likelihood Ratio," in *The Use of Statistics in Forensic Science*, ed. C. G. G. Aitken and D. A. Stoney (London: CRC Press, 1991).). A motivation is that placing the odds and *B* on a logarithmic scale permits one to think of the prior log-odds as an initial weight for H_k; a positive log-*B* adds more weight to H_k. Log-*L* also is related to expressions for information and entropy (I. .J. Good, *Good Thinking: Foundations of Probability and Its Applications* (Minneapolis, MN: University of Minnesota Press, 1983).).

[299] The most formal method of characterizing the uncertainty about the values of likelihood considers assigning a prior belief to the structure of the likelihood function (this is different than the prior belief for a proposition). Then, the likelihood for the evidence under H_k is integrated (or averaged) over all possible values, as determined by its prior distribution, to obtain the numerator and denominator of the *BF*. Because different people may choose different prior beliefs, it is expected that the value of the *BF* for the same data (evaluated by a different person) can be different. In this sense, the *BF* implicitly carries with it a greater sense of uncertainty than the *LR*.

of a quantitative *BF* (in which the values of the numerator and denominator were expressed separately and then divided), whereas the second example in box 2A.2 illustrates the use of a qualitative *BF*.[300]

The examples in box 2A.2 illustrate how both the prior odds and the *BF* can play major roles in assessing a source probability $P(H_1|e)$, and they show how a judge, juror, or other fact finder can update prior odds in light of the expert's reported *BF*.[301] This model of reasoning leads to additional argument for having the expert evaluate only the *BF* that grades the strength of the evidence. The information that affects the prior odds is outside the knowledge and expertise of the handwriting expert, who is supposed to form an opinion based only on the handwriting specimens, uncontaminated by judgements involving other evidence against the defendant. It follows that FDEs should report only the *BF* or a related expression for the weight of the evidence rather than try to judge the probability that a defendant is the source of trace evidence.

Box 2A.2: Bayes' rule in operation

According to Bayes' rule, *posterior odds* = *BF* × *prior odds*. In the case of the brother's suicide note, suppose that *BF* is 10, meaning that the FDE (correctly) believes the evidence is 10 times more probable if the surviving brother W_1 is the writer than if W_2 is the writer. If the fact finder initially believed (in light of all the other evidence about the brothers) that the odds that W_1 killed his brother were $Odds(H_1) = 2{:}1$, then the handwriting evidence changes the odds to $P(H_1|E) = BF \times Odds(H_1) = 10 \times 2{:}1 = 20{:}1$. Expressed as probabilities, the handwriting evidence has changed the probability from 2:3 (67%) to 20:21 (95%).

Now consider the case of a ransom note in Los Angeles. Suppose that *BF* is 100,000, meaning that the FDE believes that the evidence is 100,000 times more probable that W_1 is the writer than someone else (drawn at random from the city of Los Angeles). Although this *BF* is large, if the fact finder initially believed that all 4 million residents of Los Angeles were equally likely to have produced the questioned handwriting, and if this fact finder accepted the expert's estimated *BF*, then the odds of H_1 to those of H_0 would change from 1 in 4 million (before considering the handwriting evidence) to 1 in 40 (after considering the expert evidence). The corresponding subjective posterior probability assigned to H_1 would be 1:40 = 0.025, or 2.5%.

In summary, the *LR* is a measure of the evidential strength that contains a higher degree of certainty than the *BF*. However, it can be difficult to obtain the value of the *LR* for handwriting evidence. In addition, prior beliefs can be difficult to elicit, leading to use of a qualitative *BF* as a proxy for the formal *BF* or *LR*, which also contains more uncertainty than the *LR* and should be noted by the expert. Experts sometimes use a numerical scale (e.g., a 6- or 10-point scale) as a proxy for the likelihood ratio or as a more intuitive quantification of the evidential strength. FDEs can and should provide vital assistance by making explicit their use of a conventional

[300] Using a qualitative *LR* makes the resulting statistic a *BF* because it implicitly contains uncertainty regarding the exact values of the $P(E|H_k)$. That is why the first example in the box is a quantitative *BF* and the second is a qualitative *BF*.

[301] An illustrative approach may be the chart approach recommended in Kaye and Ellman 1979. (D.H. Kaye and I. M. Ellman, "Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?," *New York University Law Review* 54 (1979).). Here, the trier of fact is provided with a chart with several columns. One column lists various prior probabilities. The second lists the new information (the LR based on the test). The third is the list of various posterior probabilities. The jury members are told that it is their task—not the task of the expert—to select the prior probability. See also Ronald Meester and Marjan Sjerps, "Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence," *Law, Probability and Risk* 3, no. 1 (2004), https://doi.org/10.1093/lpr/3.1.51.

linguistic or numerical scale to express the strength of evidential support, and in their written statement and testimony they should explain how it maps onto the likelihood ratio.[302]

[302] Aitken, Roberts, and Jackson, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*.

3. Reporting and Testimony

Introduction and Scope

After the FDE completes the examination and interpretation of evidence, they still must communicate the examination results, usually by a written report or by testimony in a judicial or quasi-judicial forum. Both are important, and both must be based on sound science and reliable analytical methods.

This chapter reviews and suggests recommendations for the elements that should be part of any clear, complete report and that should be incorporated in testimony. Methods to evaluate the technical accuracy and clarity of reports and testimony are discussed, along with other means to identify and minimize the effect of human factors issues in conveying information to a client or the courts.

Different types of evidential laboratory reports exist—for example, the ENFSI guide identifies four types of reports: evaluative, technical (factual), intelligence, and investigatory.[303] Evaluative (which evaluates the forensic findings in the light of at least one pair of propositions) and technical reports (a descriptive account of observations and findings) are ordinarily used in civil and criminal cases and are the focus in this chapter.

3.1. Value of the Forensic Report

Although deposition or court testimony by the FDE is not always required, a written report may be required by laboratory accreditation bodies like the American National Standards Institute (ANSI)-ASQ National Accreditation Board (ANAB).[304] According to the accreditation program's requirements, a laboratory shall have a procedure for reporting analytical work.[305] There may be some exceptions that allow deviations from a laboratory's reporting policy.[306]

The report becomes a record of the parameters, methods, examinations, limitations, and conclusions regarding submitted evidence. For the customer, the report could point the investigation in a particular direction, inculpate or exculpate a suspect or defendant, or have a neutral impact. The report allows civil and criminal litigators to assess the evidentiary value of the examination results and may help guide the disposition of the case. Therefore, the report must be accurate, clear, and objective, detailing the analysis and comparisons of the evidence, including the conclusions and limitations. All other relevant information should be documented in the case record and available for litigant review if it is not contained in the report.

The pretrial evaluation of the report by the attorneys and investigators in the case is particularly important because many criminal and civil cases are resolved without a trial. The prosecution and the defense, plaintiff and defendant, and parties to an arbitration or administrative matter

---

[303] European Network of Forensic Science Institutes (ENFSI), *Guideline for Evaluative Reporting in Forensic Science*, Section 2.1.

[304] *ISO/IEC 17025:2005—Forensic Science Testing Laboratories Accreditation Requirements,* AR 3028, (ANSI National Accreditation Board (ANAB), 2017).Requirement 7.8.1.2.1 of the ANAB accreditation requirements makes it clear that test reports shall be provided to the customer; ISO/IEC 17025: 2017. "Shall" means "a requirement." Standard 3 Terms and Definitions.

[305] *General Requirements for the Competence of Testing and Calibration Laboratories,* ISO/IEC 17025:2017, AR 3125, (ANSI-ASQ National Accreditation Board (ANAB), 2017). Requirement 7.8.1.2.2

[306] There may be differences in reporting requirements between civil and criminal cases (see section 3.4). In addition, private practitioners may not be subject to the same guidelines as accredited laboratories.

must evaluate the significance of the report's conclusion and determine the weight to give it in plea and settlement discussions. The laboratory report informs the parties on crucial strategic decisions. The pretrial examination of the report is where the contents and structure of the report, described in section 3.4.1, become important for understanding the influence of the forensic examination in the case.

In addition to pretrial use, the report may serve as a stand-alone document during court proceedings without testimonial support by the FDE.[307] If there is a stipulation between the parties regarding the findings and conclusions of the expert, the report may be read to the jury and put into the court record.[308] In such cases, the report alone must accurately represent the bases of the FDE's findings and conclusions.

In court, the laboratory report is the foundation of the FDE's testimony, whether evaluative or technical, and the FDE must be able to decipher, clarify, explain, and defend its contents to the fact finder. The FDE must possess a working knowledge of the discipline, be able to explain the foundational principles of handwriting analysis and the fundamentals of the discipline's validity and reliability (including studies supporting those concepts), and be familiar with the studies indicating potential or known error rates. Visual aids used to educate the jury and explain the FDE's conclusions must be prepared and presented in an unbiased manner consistent with the report and the anticipated testimony.

3.2. The Forensic Report and Human Factors

A comprehensive report not only includes the necessary technical content but also clearly conveys that information to the report's recipients. International Organization for Standardization (ISO) guidelines, for example, require each test to be reported "accurately, clearly, unambiguously and objectively."[309] This standard has been adopted by forensic science laboratory accreditation bodies.[310] When preparing a report and translating the processes and conclusions into plain, understandable language, human factors must be considered. The author's educational background, professional training, attitude toward the job, and cognitive biases among other human factors all affect the report's content and form. Writing the report reflects on the methods of analysis and evaluation and anticipates future direct and cross-examination.

---

[307] Despite the prohibition in *Melendez-Diaz v. Massachusetts, 557 U.S. 305 (2009).* that barred the introduction of a laboratory report without the ability of the defendant to confront the analyst, there remain constitutionally valid "notice-and-demand" statutes in some states by which the prosecution provides the defendant with notice of its intent to introduce the laboratory report without calling the analyst. The defendant can then assert their right to have the analyst present in court to testify or forfeit that right by silence. *Id.* at 326 and cases cited therein.

[308] For example, in Melendez-Diaz v. Massachusetts, U.S., Justice Scalia noted that in drug cases "[d]efense attorneys and their clients will often stipulate to the nature of the substance in the ordinary drug case." At least in Massachusetts, it is "'almost always the case that [analysts' certificates] are admitted without objection.'" *Id.* at 328.

[309] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Section 7.8.1.2. ISO, a non-government international organization, creates voluntary, consensus-based international standards. ISO has partnered with its sister organization, IEC, which sets consensus-based international standards for electrical, electronic, and related technologies. Together, they have published standards for the competence of testing and calibration laboratories. The version current at the time of this report's publication is known as ISO/IEC 17025:2017.

[310] Another international standard for assessment of forensic science service providers is ISO/IEC 17020:2012. That standard is most often used for crime scene investigation units. The standards for contents of inspection reports contained in Section 7.4 and Appendix B are not as robust as those contained in ISO/IEC 17025. Elements of the inspection reports found in Appendix B are optional. Examples of the optional information include information on what has been omitted from the original scope of work; identification or brief description of the inspection method(s) and procedure(s) used, mentioning the deviations from, additions to, or exclusions from the agreed methods and procedures; and identification of equipment used for measuring/testing.

There may be fewer human factors involved when writing a simple, skeletal laboratory report like that in *Melendez-Diaz v. Massachusetts*, which read in its entirety "the substance was found to contain: Cocaine"[311] (however, many human factors may have played a role in the analysis underlying the report). Today, the narrative portion of a laboratory report is often a more comprehensive document, telling a story in the life of a piece of evidence. The narrative might describe the documentary evidence, where it came from (chain of custody), why it is to be examined, how it was examined, and the conclusion or opinion derived from its examination.

A laboratory report must be understandable and have a logical flow for its conclusions to have meaning. It should account for all data, both pro and con the proposition, and for alternative propositions. Because "[f]*orensic reports* are instances of communicative behavior written about specific [evidence] and for audiences with specific needs,"[312] the experiences of both author and reader play a role. Initially, the cognitive biases of the author must be mitigated by robust laboratory procedures or other means. For example, if known evidence is examined before reviewing questioned evidence, this sequence should be reflected in the report, so that any reader of the report is alerted to the potential for cognitive bias (see the process map [figure 1.1] and section 2.1). The challenge is not to import new biases as the data are reviewed. The author should question every assertion made in the report and consider everything done in the examination to increase the report's utility and avoid error. Transparency in the analytical and evaluative processes allows internal laboratory reviews and critical external assessments by criminal justice stakeholders to be more effective, which, in turn, allows a greater opportunity to detect errors.

The act of writing the report can have cognitive effects on the FDE.[313] Language communicates the FDE's work and conclusions, and the formulation of the language can affect the FDE's cognition. By focusing on validity, reliability, and objectivity, the FDE can remain as impartial as possible when writing the report, rather than taking on the inappropriate role of advocate.

Cognitive issues must also be considered for those who read the report. Each party in the litigation, each judge and each juror, has pre-existing personal biases. In addition, criminal and civil cases may introduce cognitive issues affecting the reader's interpretation of the report like the case facts, confirmation or expectation bias, framing, and advocacy blinders, which may affect how the reader understands the conclusion. The FDE's challenge is to write the report in a way that mitigates those cognitive factors by writing a clear, unambiguous report based on an established scientific examination method.

Language also affects how information is perceived by the reader. Neumann and Reyna state that "[j]urors have a poor understanding of the terms conventionally used to report the conclusions of forensic examinations and are generally confused by conclusions reported using

[311] Melendez-Diaz v. Massachusetts, U.S.

[312] Michael Karson and Lavita Nadkarni, *Principles of Forensic Report Writing* (2013), 11.

[313] Itiel E. Dror, "Cognitive Neuroscience in Forensic Science: Understanding and Utilizing the Human Element," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370, no. 1674 (Aug 5 2015), https://dx.doi.org/10.1098/rstb.2014.0255, https://www.ncbi.nlm.nih.gov/pubmed/26101281.

probabilities."[314] As such, the FDE needs to be cognizant of how the language and descriptions in the report can aid or hinder a naive reader.

Furthermore, all readers may not interpret the meaning and consequences of information in the same way or in the way that the FDE intended. Neumann and Reyna[315] give examples of human factors affecting an individual's perception of what is reported about a latent print identification and a fiber transfer. They assert that the impact of a conclusion

> can vary depending on personal experience, background, knowledge of transfer of trace material in similar situations, education about the respective probative value of fingerprint and fiber evidence, and general importance of the evidence in the case. The consequences for the defendant, in terms of support for innocence or guilt and associated sentence, can also affect the interpretation of the statement.[316]

These variables may likewise impact the perceptions of a handwriting examination report. Jurors might also be influenced by their evaluation of the FDE's experience. One conclusion from an NIJ report[317] stated

> The findings suggest that jurors tend to over-value some attributes of forensic science expert testimony and under-value other aspects. The most persistent finding is that jurors rely heavily on the "experience" of the testifying expert and the expert's asserted certainty in his conclusions.

This is troubling for two reasons. First, research has shown that accuracy in handwriting examination determinations is not related to years of experience.[318] Second, jurors (and presumably other consumers of forensic reports or testimony) tend to prefer certainty. Jackson and Roesch[319] report on two studies in this regard.

> Another way in which researchers have studied expert certainty is to manipulate the extent to which the expert's conclusions are unambiguous in favoring one side of the case, or are more cautious or balanced in acknowledging possible limitations. The two studies that have manipulated this aspect of certainty indicate that jurors prefer unambiguous testimony that is strongly worded. For example, Brekke, Enko, Clavet, and Seelau [citation omitted] manipulated whether the testimony was slated in favor of the prosecution or balanced. In the balanced conditions, the expert discussed limitations of the evidence. Results indicated that, as expected, the slanted testimony yielded the highest conviction rates for dependence in both the prosecution and court-appointed expert conditions. The slanted testimony was all rated as being more useful and of higher quality than the more balanced testimony that acknowledged the presence of some shortcomings. Rudy [citation omitted] manipulated the strength of the expert's testimony in a sexual abuse case. There were no significant differences in verdict

---

[314] C. Neumann and V. Reyna, *Jury Studies and the Psychology of Efficient Communication*, NIST (11/8/15 2015), 32, https://www.researchgate.net/publication/291817447_Communicating_the_Results_of_Forensic_Science_Examinations.

[315] Neumann and Reyna, *Jury Studies and the Psychology of Efficient Communication*, 35.

[316] Neumann and Reyna, *Jury Studies and the Psychology of Efficient Communication*.

[317] N. J. Schweitzer, *Communicating Forensic Science*, National Criminal Justice Reference Service (2016), 10–11, https://www.ncjrs.gov/pdffiles1/nij/grants/249804.pdf.

[318] Sita, Found, and Rogers, "Forensic Handwriting Examiners' Expertise for Signature Comparison," 1117, 23.

[319] R. Jackson and R. Roesch, eds., *Learning Forensic Assessment: Research and Practice*, 2nd ed. (New York: Routledge, 2016), 516.

> between jurors hearing a high-certainty expert statement and more neutral testimony. However, mock jurors rated the high-certainty testimony as more credible than the neutral testimony.

These findings are concerning, because if jurors and others give greater credence to strong opinions that might not be as well-reasoned or well-founded as more complex, qualified opinions, they may make incorrect decisions on culpability or liability. FDEs should not push their opinions to stronger levels of confidence than merited by the evidence to convince jurors; instead, experts should thoroughly explain the reasons for qualifications and the importance of limitations.

Other factors may also affect the weight that fact finders give to the testimony of experts and the probative value of their conclusions. One factor is the presentation format for the conclusion like a numerical versus verbal expression of the likelihood ratio.[320] When using random match probabilities like in DNA analyses, other factors include the "prosecution fallacy," an "assumption that the random match probability is the same as the probability that the defendant was not the source of the DNA sample..."[321] and the "defense fallacy," which "resembles the prosecutor's fallacy in making an illogical leap but differs in understating the tendency of a reported match to strengthen source probability and narrow the group of potential suspects."[322] The introduction of false report probabilities (false positives) also may create the possibility of errors in the assessment of forensic evidence, called the "false positive fallacy."[323]

In a 2015 article by Dror et al.,[324] the authors discuss jury instructions from judges in cases where there is concern over cognitive bias from the experts. In part, that section reads

> … courts should consider giving a jury instruction regarding cognitive bias and the risk factors that may affect an expert's judgment and conclusion. This is already somewhat common in eyewitness identification cases where jury instructions on how memory works are now regularly given. There is ample science to support an instruction for evaluating expert cognitive bias.

Although it would be helpful if judges would also instruct the jury about the potentially equal or superior strength of qualified and inconclusive opinions over unqualified opinions, this is in the province of the court. What the FDE can and should do is make it clear in the report or testimony that "inconclusive," "no conclusion," "insufficient for examination," "qualified opinions," and

---

[320] K. A. Martire et al., "The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect," *Law and Human Behavior* 37, no. 3 (Jun 2013), https://doi.org/10.1037/lhb0000027.

[321] *State v. Small*, 184 A.3d 816, 825 (Conn.App. 2018).

[322] *United States v. Chischilly*, 30 F.3d 1144 (9th Cir. 2014). Also see William C. Thompson and Edward L. Schumann, "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy," *Law and Human Behavior* 11, no. 3 (1987): 359, 62–64, https://doi.org/10.1007/bf01044641; William C. Thompson, Suzanne O. Kaasa, and Tiamoyo Peterson, "Do Jurors Give Appropriate Weight to Forensic Identification Evidence?," *Journal of Empirical Legal Studies* 10, no. 2 (2013), https://doi.org/10.1111/jels.12013.

[323] Thompson, Kaasa, and Peterson, "Do Jurors Give Appropriate Weight to Forensic Identification Evidence?," 359, 62–64.

[324] Itiel E. Dror, B. M. McCormack, and J. Epstein, "Cognitive Bias and its Impact on Expert Witnesses and the Court," *The Judges' Journal* 54, no. 4 (2015): 359, 62–64, https://www.americanbar.org/groups/judicial/publications/judges_journal/2015/fall/cognitive_bias_and_its_impact_on_expert_witnesses_and_the_court/.

"unqualified opinions" can all be equally valid, explanatory, and meritorious opinions and should therefore be viewed by the consumer of the report as informative statements.

Dror[325] argues that "most people view reporting in a cognitively naïve way, i.e., that the report simply reflects the working of the forensic examiner." As noted previously, the report is much more than a reflection of the analysis or the opinion of the examiner.

3.3. Opinion Scales

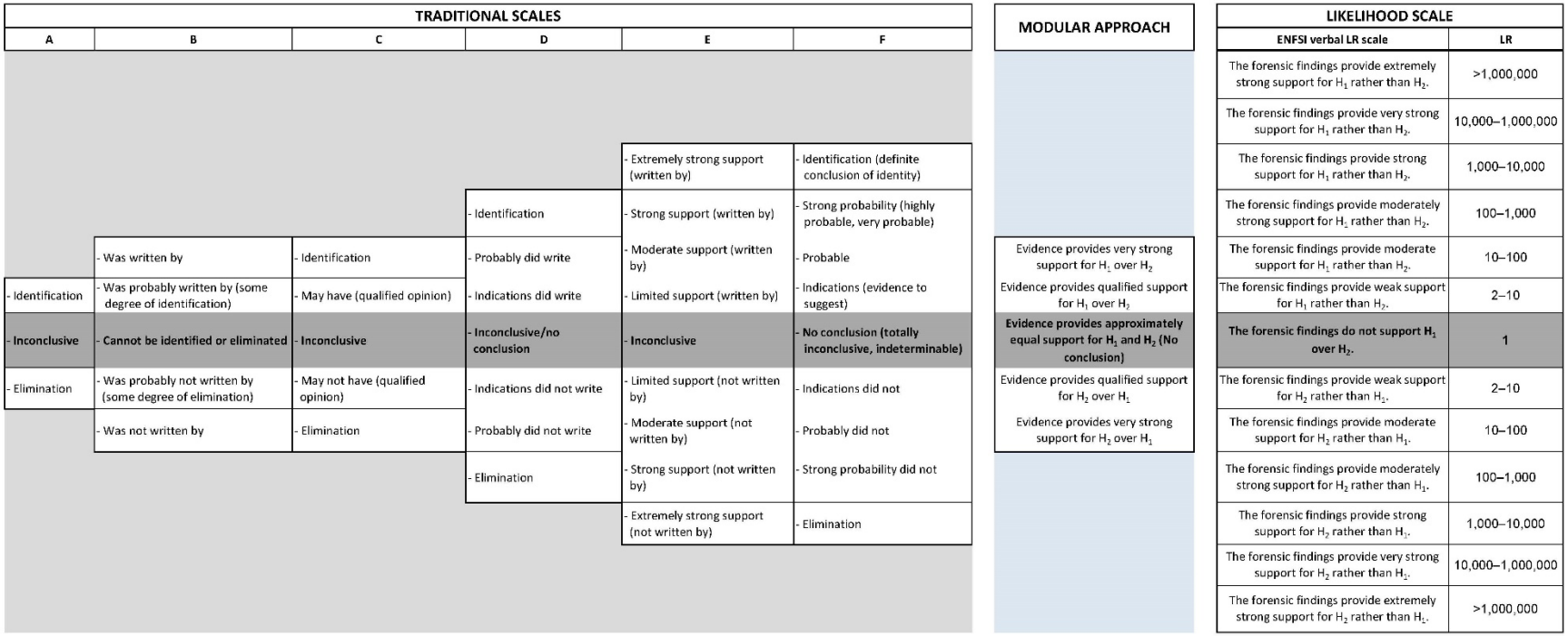
Figure 3.1 presents examples of the different sets of conclusion terms used globally in forensic handwriting examination. These terms are generally referred to as opinion scales. Although opinion scales are not scientifically rigorous, FDEs and the courts often view conclusion terminology as ordinal (or strength) scales. This view has some inherent problems. An ordinal scale arises from the function of rank ordering[326] and can demonstrate a gradation of strength of the FDE's opinion. However, the level of gradations between the opinion levels are not quantified (except in the likelihood ratio scale). For example, it is not possible for an FDE to define clearly the degree of difference between "highly probable" and "probable." All the FDE can say is that probable is the weaker or less strong of the two opinion levels. There may be variance among FDEs in how they view the degree of difference between the opinion levels.

In the conventional set of scales (five, seven, or nine points), the FDE expresses opinions corresponding to the conventional approach to handwriting analysis (see section 1.3). Although these opinions may be stated in probabilistic terms (e.g., probably wrote), their precise meaning may be inconsistent across FDEs. For example, some FDEs may render an opinion based on the rarity of features and others based on perceived evidential strength. When presenting evidence using the conventional scales, there is always a step where the FDE decides whether or not the writer of the known writing samples could have written the questioned document. In contrast, when using the modular[327] and likelihood ratio–based approaches (see section 2.3.2), the FDE is expressing the strength of the evidence in terms of two or more mutually exclusive propositions or hypotheses without first considering the typicality of the questioned document given what is known about the suspected writer. This is generally expressed as the strength of support for one proposition or hypothesis over one or more mutually competing propositions.

[325] Dror, "Cognitive Neuroscience in Forensic Science: Understanding and Utilizing the Human Element," 3.

[326] S. S. Stevens, "On the Theory of Scales of Measurement," *Science* 103, no. 2684 (Jun 7 1946), https://doi.org/10.1126/science.103.2684.677.

[327] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83.



TRADITIONAL SCALES						MODULAR APPROACH	LIKELIHOOD SCALE	
A	B	C	D	E	F		ENFSI verbal LR scale	LR
				- Extremely strong support (written by)	- Identification (definite conclusion of identity)		The forensic findings provide extremely strong support for H_1 rather than H_2.	>1,000,000
			- Identification	- Strong support (written by)	- Strong probability (highly probable, very probable)		The forensic findings provide very strong support for H_1 rather than H_2.	10,000–1,000,000
	- Was written by	- Identification	- Probably did write	- Moderate support (written by)	- Probable	Evidence provides very strong support for H_1 over H_2	The forensic findings provide strong support for H_1 rather than H_2.	1,000–10,000
- Identification	- Was probably written by (some degree of identification)	- May have (qualified opinion)	- Indications did write	- Limited support (written by)	- Indications (evidence to suggest)	Evidence provides qualified support for H_1 over H_2	The forensic findings provide moderately strong support for H_1 rather than H_2.	100–1,000
- Inconclusive	**- Cannot be identified or eliminated**	**- Inconclusive**	**- Inconclusive/no conclusion**	**- Inconclusive**	**- No conclusion (totally inconclusive, indeterminable)**	**Evidence provides approximately equal support for H_1 and H_2 (No conclusion)**	The forensic findings provide moderate support for H_1 rather than H_2.	10–100
- Elimination	- Was probably not written by (some degree of elimination)	- May not have (qualified opinion)	- Indications did not write	- Limited support (not written by)	- Indications did not	Evidence provides qualified support for H_2 over H_1	The forensic findings provide weak support for H_1 rather than H_2.	2–10
	- Was not written by	- Elimination	- Probably did not write	- Moderate support (not written by)	- Probably did not	Evidence provides very strong support for H_2 over H_1	**The forensic findings do not support H_1 over H_2.**	**1**
			- Elimination	- Strong support (not written by)	- Strong probability did not		The forensic findings provide weak support for H_2 rather than H_1.	2–10
				- Extremely strong support (not written by)	- Elimination		The forensic findings provide moderate support for H_2 rather than H_1.	10–100
							The forensic findings provide moderately strong support for H_2 rather than H_1.	100–1,000
							The forensic findings provide strong support for H_2 rather than H_1.	1,000–10,000
							The forensic findings provide very strong support for H_2 rather than H_1.	10,000–1,000,000
							The forensic findings provide extremely strong support for H_2 rather than H_1.	>1,000,000

*The depiction of the different scales adjacent to each other is not meant to demonstrate a one-to-one mapping or show direct correlation between the scales but rather to illustrate the different opinions most commonly employed by FDEs.

A = Conclusions that are often required by handwriting studies

B = 5-point opinions used by Collaborative Testing Services (CTS)

C = 5-point opinions used by the Federal Bureau of Investigation (FBI)

D = 7-point opinions

E = 9-point opinions defined by the European Network of Forensic Handwriting Experts (ENFHEX)) in their Collaborative Exercise program

F = 9-point opinions outlined by the Scientific Working Group for Forensic Document Examination (SWGDOC)

Figure 3.1: Presentation of conventional conclusions and the likelihood-based scale*

The three levels that FDEs currently use that are present consistently across the scales are identification, inconclusive, and elimination. In the modular approach, there are no identification or elimination opinions. Currently, there is no way to map or relate the different types of scales for the following reasons:

1. The conventional scales address the probability of the proposition whereas the modular and likelihood ratio approaches focus on the probability of the findings given the proposition. As such, the conventional scales cannot be equated to the other approaches.
2. All scales lack sufficient study and empirical evaluation; therefore the consistency of application across FDEs is not well understood.
3. There would be fundamental mathematical issues in attempting to map the discrete categories in the different scales unless there was some common reference point or "anchor" between each scale.

The definitive conclusions (identification and elimination) on all of the conventional scales appear to have consistent application across the FDE community. The scales also share the center point but not the range of the inconclusive category. Although the different scales might share the same meaning for identification, elimination, or possibly inconclusive, the sufficiency of evidence that an individual FDE may use to support that conclusion may not be equivalent.

FDEs have reported[328] that the actual category boundaries of the scale are subjectively determined during their evaluation, depending on the extent of perceived differences or similarities among the questioned and known writings and limitations of the materials examined. For example, the decision matrix for the nine-point scale reporting conclusions suggests that a finding of Identification should be made if the "range of variation in the questioned writing and in the known writing contains substantial significant [i.e., relevant] similarities" and there are "no significant dissimilarities," whereas a finding of Indications Did Write should be reported if the "range of variation exhibited in the questioned writing and in the known writing contains few significant similarities" and there are "no significant dissimilarities."[329] The difference between a few and substantial similarities is undefined.

In a black box study, one of the measures is consistency among FDEs when evaluating a given sample set. However, these studies must take the variety of conclusion scales into account; otherwise, if FDEs are unfamiliar with the particular conclusion scale used in a given study, it may lead to study findings that are not reflective of actual casework and may be of little value in moving the field forward.

To begin moving toward a unified, standard approach for expressing conclusions, the FDE community should address some of the issues above by taking some bold, albeit difficult, actions like the following:

[328] Merlino et al., *Validity, Reliability, Accuracy, and Bias in Forensic Signature Identification*.

[329] Scientific Working Group for Forensic Document Examination (SWGDOC), "Guidelines for Forensic Document Examination," *FBI Forensic Science Communications* 2, no. 2 (2000).

- Begin using uniform conclusion scales that explicitly describe the propositions considered;
- Create a uniform training set with ground-truth-known answers and a consensus for the appropriate conclusion based on the limitations of the evidence, in the context of a multiple proposition method;
- Train all new FDEs across the community using the same dataset and with uniform tests; and
- Retrain existing FDEs, to the extent required, to have a working knowledge of the conclusion sets using a dual-proposition method in a transparent manner.

3.4. The Forensic Report on Handwriting Examinations

The Working Group began its analysis of the content and format of FDE reports by reviewing extant legal and accreditation requirements and recommendations from other relevant groups. Best practices from these materials and from practitioners in the forensic handwriting examination community were compiled and analyzed, resulting in recommendations by the Working Group (see recommendations 3.1 and 3.2).

Communication is a critical human factors issue, and the forensic report often serves as a primary means of communication between the scientist and others within the criminal justice community. Discussions of report content should incorporate aspects that affect human factors issues within the context of the designated requirements. However, before discussing report content, it is important to review the requirement for the FDE to prepare a report. For instance, the Federal Rules of Criminal (Rule 16) and Civil (Rule 26) Procedure treat the requirement of written reports, otherwise known as court statements, differently. Although these rules only govern the federal courts, many state courts model their rules after them. It makes sense, then, that forensic science reports contain at a minimum the information required by the rules of discovery, if for no other reasons than for the efficiency of the expert and as an accommodation for the customers' litigation responsibilities. The following paragraphs reflect the Working Group's understanding of relevant requirements and case law, and the Working Group acknowledges that others may interpret the referenced subject matter differently.

The Civil Rule requires that when disclosure of expert testimony is made, the "disclosure must be accompanied by a written report—prepared and signed by the witness—if the witness is one retained or specially employed to provide expert testimony in the case."[330]

On the other hand, the Criminal Rule only requires each side to provide an opportunity to "inspect and to copy or photograph the results or reports of any physical or mental examination and of any scientific test or experiment"[331] (emphasis added). The NCFS recommended—both as a matter of fairness and to promote the accurate determination of the truth—that prosecutors keep pretrial disclosure of forensic science reports in line with "the federal civil rules presently

[330] Federal Rules of Civil Procedure Rule 26(a)(2)(B).

[331] Federal Rules of Criminal Procedure Rule 16(a)(1)(F).

require than the more minimal requirements of the federal criminal rules."[332] The Working Group agrees with that recommendation.

Anecdotally, it has been noted that some attorneys fail to ask for a written report from FDEs or ask them not to write a report, thereby avoiding some discovery obligations. Federal courts have ruled that Rule 16(a)(1)(F) and 16(b)(1)(B) require the prosecution and defendant to disclose the results or reports of any scientific test or experiment. The 1993 amendments to Rule 16 added the requirement to disclose a written summary of the expert's opinions, bases, and reasons for those opinions and the witness's qualifications. That amendment solved the problem of non-disclosure of oral reports, because a summary of the testimony must be provided even for oral reports.[333]

When the FDE is employed by an accredited laboratory, however, a written or electronic report is likely required each time an examination is conducted. According to the ANAB accreditation requirements, a laboratory shall have a procedure for reporting results that, among other things, "identifies what will be reported for all items received, including items on which no work was performed, items collected or created and preserved for future testing, and for all (partial and complete) work performed."[334]

Even though written reports are expected when an analysis has been conducted in an accredited laboratory, in some exigent criminal and national security cases FDEs may be asked to make oral or preliminary reports as investigatory leads. These reports are sometimes referred to as intel reports and can deviate from QA policies like technical review requirements. When such reports are issued, FDEs should document the examinations in the case records and prepare reports subject to the QA procedures expressing the limitations of the examinations and conclusions for later disclosure pursuant to legal requirements. Appropriate limitations in examination and conclusions should be stated, along with a statement that any conclusion may change with a full examination. FDEs should also be aware of the enhanced danger of cognitive bias and the potential for reduced reliability because of the real possibility that task-irrelevant information will be communicated by the investigator to the FDE as part of emerging facts in an ongoing investigation; such concerns should also be communicated to the readers of the report. If the examined evidence will be the subject of expert testimony in court, the evidence should be re-examined by another FDE and a new report prepared.

Unlike accredited laboratories, those FDEs whose laboratories are not accredited may not be required to write a report each time an analysis is conducted, but the analyses and conclusions should be documented in the FDE's case record. The particular legal situation and status of the FDE may also influence whether a report is written. For example, a consulting expert for a civil litigant or a criminal defendant does not have to disclose the results of the analysis to the

[332] National Commission on Forensic Science (NCFS), *Recommendations to the Attorney General: Pretrial Discovery*, Department of Justice (2016), https://www.justice.gov/ncfs/file/880241/download.

[333] See, for example, *United States v. Smith*, 101 F.3d 202 (1st Cir. 1996); *United States v. Shue*, 766 F.2d 1122 (7th Cir. 1985).

[334] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Requirement 7.8.1.2.2

opposing party unless and until the FDE is identified as a testifying expert and then only pursuant to the court's discovery rules.[335]

Recommendation 3.1: Whenever a handwriting examination is conducted, forensic document examiners should prepare reports as described in Recommendation 3.2, unless exempt by documented laboratory policy.

3.4.1. Contents of the Forensic Report

A baseline for report content is found in the same Federal Rules of Criminal (Rule 16) and Civil (Rule 26) Procedures that provide for advance disclosure of the nature and basis of expert testimony expected to be proffered under FREs 702, 703, or 705. To the extent that the rules specify the nature of the information to be disclosed in discovery, they shed light on what the Advisory Committees on the Federal Rules of Criminal Procedure and Civil Procedure believe is necessary to avoid surprise and to provide an opportunity for the opponent to "test the merit of the expert's testimony through focused cross-examination"[336] and to arrange for expert testimony from other witnesses.[337] Advance disclosure also allows the opponent to move for a pretrial hearing on the admissibility of the expected expert testimony (e.g., a *Daubert*[338] hearing), to obtain additional testing, and to find a rebuttal expert.

The civil discovery rule requires a written report that must contain a complete statement of all opinions the witness will express and the bases and reasons for them. In addition, the report must contain the facts or data considered by the expert in forming the opinions and all supporting exhibits. This provision is to be broadly interpreted and requires not only disclosure of the facts or data relied upon to arrive at the conclusions or opinions but also those merely considered by the expert.

The criminal discovery rule, however, requires only a written summary that describes the expert's opinions and the bases and reasons for those opinions. That summary, according to the Advisory Committee Notes, should include "any information that might be recognized as a legitimate basis for an opinion under Federal Rule of Evidence 703."[339]

The NCFS[340] recommended to the Attorney General that the report provided in discovery should contain

> (i) a statement of all opinions the witness will express and the basis and reasons for them; (ii) the facts or data considered by the witness in forming them; (iii) any exhibits that will be used to summarize or support them; (iv) the witness's qualifications, including a list of all publications authored in the previous 10 years; (v) a list of all other cases in which, during

[335] Federal Rule of Criminal Procedure 16(b)(1)(B); Federal Rule of Criminal Procedure 16(b)(1)(C); Federal Rules of Civil Procedure 26(b)(4)(D); *United States v. Walker*, 910 F.Supp. 861 (N.D.N.Y. 1995).

[336] Federal Rules of Criminal Procedure Rule 16.

[337] Advisory Committee Notes, *Advisory Committee on Rules of Civil Procedure—May 1993* (Washington, DC, 1993), 26.

[338] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.

[339] F.R.E. 703. *Bases of an Expert's Opinion Testimony*, says in part: "An expert may base an opinion on facts or data in the case that the expert has been made aware of or personally observed. If experts in the particular field would reasonably rely on those kinds of facts or data in forming an opinion on the subject, they need not be admissible for the opinion to be admitted."

[340] National Commission on Forensic Science (NCFS), *Recommendations to the Attorney General: Pretrial Discovery*.

the previous 4 years, the witness testified as an expert at trial or by deposition; and (vi) a statement of the compensation to be paid the witness.

The requirement to disclose the bases and reasons for the expert's opinions is consistent with the Advisory Committees' emphasis on focused cross-examination of the expert. The U.S. Supreme Court agreed in 1993, stating in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* that "vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof" is not only the conventional method, but also an appropriate means to attack "shaky but admissible evidence."[341] Sixteen years later, the Supreme Court again stressed the importance of cross-examination of expert witnesses. In *Melendez-Diaz v. Massachusetts*, Justice Antonin Scalia argued that "there is little reason to believe that confrontation will be useless in testing analysts' honesty, proficiency, and methodology—the features that are commonly the focus in the cross-examination of experts."[342] The high court's trust in cross-examination reaffirms the need for forensic scientists to write reports that give opponents fair notice of the tests performed and the opinions reached by experts.

The NCFS Reporting and Testimony Subcommittee characterized the functional equivalent of "peer review" within the legal system to be the examination and cross-examination of proffered scientific evidence. Advance disclosure through the discovery process should include the "kinds of analyses conducted and methods used to evaluate those items; the testing conducted on those items; the observations made; the opinions, interpretations, and conclusions reached; and the bases for those observations, opinions, interpretations, and conclusions."[343]

The importance of complete test reports is highlighted by the application of the FRE, primarily FRE 702. Modified in 2000 in response to the *Daubert* trilogy,[344] FRE 702 sets the stage for the admissibility of expert testimony, including that which is scientific, technical, or based on specialized knowledge. Although *Daubert*'s non-exclusive considerations for assessing the validity and reliability of expert testimony are discretionary with a court, FRE 702 sets forth four general factors that federal courts, and some state courts that have adopted FRE 702, use in assessing admissibility. Rule 702[345] states the following:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

1. The expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
2. The testimony is based on sufficient facts or data;
3. The testimony is the product of reliable principles and methods; and

[341] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.

[342] Melendez-Diaz v. Massachusetts, U.S.

[343] National Commission on Forensic Science (NCFS), *Views of the Commission: Pretrial Discovery of Forensic Materials*, Department of Justice (2015), https://www.justice.gov/ncfs/file/786611/download.

[344] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S; *General Electric Co. v. Joiner*, 522 U.S. 136 (1997); *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).

[345] F.R.E. 702. Testimony by Expert Witnesses

4. The expert has reliably applied the principles and methods to the facts of the case.

The application of FRE 702 may begin with a motion by the opponent requesting the court, pursuant to FRE 104(a), to determine the preliminary question of whether the evidence is admissible. In response to such a motion, the proponent of the evidence is required to prove by a preponderance of evidence that the proffered testimony is admissible under FRE 702.[346] The process to accomplish that goal may be a *Daubert* hearing, or what some courts call a *Kumho*[347] hearing, depending on the nature of the evidence or the opposition to it.

The role of discovery and the completeness of test reports are important preconditions to this process. The Advisory Committee Notes for Rule 16 suggest that the basis for providing a summary of the expected testimony is to "permit more complete pretrial preparation by the requesting party."[348] Thus, counsel opposing the introduction of forensic evidence can better evaluate the need for a pretrial hearing if a full disclosure of the scientific methodology, conclusions, opinions, limitations, and bases are revealed so they can be reviewed by the opponent or the opponent's expert.

A chemist's generic test report, for example, does not meet the requirements of Rule 16 if it does not address these issues and only describes the substance found and its weight, along with a summary of the bases for the conclusions being the FDE's training, formal education, and experience, including conducting numerous drug tests. The Sixth Circuit Court of Appeals held in *United States v. Davis* that the prosecution did not meet the requirements of the rule, concluding that the defendant's chemist, if he had hired one, "would not have been able to analyze the steps that led the government's chemists to their conclusions."[349] The court also opined that it was proper for the district court to request that the chemists provide their notes to defendant's counsel.

Forensic laboratories and FDEs should recognize the importance of providing test reports that disclose methods, protocols, and standards for purposes of cross-examination. The critique inherent in cross-examination can provide useful feedback to the FDE and the forensic science community and is one way in which continuous improvement can be achieved.

Guidelines from various forensic science–related entities informed the Working Group's suggestions for report writing in handwriting examinations. Although these organizations do not focus directly on the impact of human factors in report writing, many of the guidelines account for the influence of human factors that the Working Group has recognized. These accreditation bodies are recognized by international organizations to conduct conformity assessments of forensic science service providers in compliance with ISO/IEC 17025.[350]

---

[346] See *Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S; Bourjaily v. United States, 483 U.S. 171 (1987).*

[347] Kumho Tire Co. v. Carmichael, U.S. The *Kumho* hearing is one in which the reliability or application of the method of analysis at hand is questioned.

[348] Federal Rules of Criminal Procedure Rule 16.

[349] *United States v. Davis*, 514 F.3d 596 (6th Cir. 2008).

[350] For example, ANAB is a signatory of the International Laboratory Accreditation Cooperation (ILAC) multilateral recognition arrangement (MRA). See https://www.anab.org/about-anab and https://ilac.org/signatory-search/. ILAC states that it is the international organization for accreditation bodies operating in accordance with ISO/IEC 17011 and involved in the accreditation of conformity assessment bodies, including testing laboratories (using ISO/IEC 17025). Accreditation of conformity assessment bodies, according to ILAC, is the independent evaluation of accreditation organizations against recognized standards to carry out specific activities to ensure their impartiality and competence. The ILAC

ISO/IEC 17025:2017, section 7.8.1.2, establishes an overall standard for report writing. Test results "shall be provided accurately, clearly, unambiguously and objectively, usually in a report (e.g. a test report or a calibration certificate or report of sampling), and shall include all the information agreed with the customer and necessary for the interpretation of the results and all information required by the method used. All issued reports shall be retained as technical records." In addition to identifying information and chain-of-custody authentication, the standard requires documentation for the bases and interpretations appearing in the report.[351] Opinions and interpretations in the report are to be clearly marked as such.[352] Information not included in the report must be readily available in the laboratory file.[353]

Whereas ISO establishes the international standards for laboratory competency to carry out tests and/or calibrations, the International Laboratory Accreditation Cooperation (ILAC) is an international authority that provides the infrastructure to support the exhibition of competence worldwide through accreditation programs. ILAC-G19:08/2014, *Modules in a Forensic Science Process* (hereafter ILAC-G19) was published to provide guidance for forensic units in applying ISO/IEC 17025 and ISO/IEC 17020. Section 4.9 of ILAC-G19 dictates that all reports shall meet the reporting requirements of the ISO standards.

ILAC-G19 also provides some flexibility for how the required information is conveyed, depending on legislation controlling the particular forum. Alternate ways of disclosing the report's information may be to include all the ISO/IEC 17025 information in the report, to prepare an annex to the report containing the additionally required information, or to ensure that the pertinent case record contains all the relevant information.[354] A case record includes all information relating to the analysis and would include a "technical record" that would allow "another reviewer possessing the relevant knowledge, skills, and abilities [to] evaluate what was done and interpret the data." [355]

The NCFS also recognized that a forensic report may contain less information than is present in a full case record. The NCFS suggested that the report contain the following statement: "This report does not contain all of the information needed to independently evaluate the work performed or independently interpret the data. Such an evaluation requires a review of the case record."[356]

website indicates the accreditation bodies that are signatories to the ILAC MRA have been peer evaluated in accordance with the requirements of ISO/IEC 17011 to demonstrate their competence to conduct conformity assessments. The ILAC multilateral recognition arrangement signatories then assess and accredit conformity assessment bodies according to the relevant international standards including testing laboratories (using ISO/IEC 17025). See https://ilac.org/. The A2LA is also a signatory to the ILAC MRA.

[351] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Sections 7.8.2.1 and 7.8.7.1.

[352] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Section 7.8.7.2

[353] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Section 7.8.7.3

[354] *Modules in a Forensic Science Process,* ILAC-G19:08/2014, (International Laboratory Accreditation Cooperation (ILAC), 2014). Section 4.9

[355] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Section 7.5.1.3

[356] National Commission on Forensic Science (NCFS), *Views of the Commission: Documentation, Case Record and Report Contents*, Department of Justice (2015), https://www.justice.gov/ncfs/file/818191/download.

Regardless of how the totality of information is made available, ISO/IEC 17025:2017 makes clear that in all cases, the report shall indicate which parts are background information, which are facts, and which are interpretations or opinions.

The ILAC-G19 Guidelines[357] regarding a report also specify that:

> The output given to the customer shall not in any way be misleading.
>
> The report should contain all the results of examinations/tests and observations as well as the findings and, where appropriate and admissible, conclusions drawn from these results.
>
> The reports issued by the forensic unit shall be complete and shall contain the information on which an interpretation might be made.
>
> Conclusions shall be properly qualified.
>
> It shall be clear in the report to the customer on what an interpretation and/or conclusion is based, including the results and findings, also the available information at the time of the evaluation presented in the report.

Accreditation bodies that assess forensic laboratories in light of ISO/IEC 17025 must follow those test report standards and the implementation guidance provided by ILAC but may also add supplemental accreditation requirements for report writing. Three of North America's accreditation programs for forensic laboratories are ANAB and American Association for Laboratory Accreditation (A2LA) (both ILAC signatories) and the Standards Council of Canada.[358] They assess laboratories in conformance with ISO/IEC 17025 standards, enhancing uniformity throughout the forensic science community.

When opinions or conclusions are reached that involve associations between evidentiary items, the ANAB program accreditation requirements direct that the significance of an association must be communicated clearly and qualified properly in the test report. The reasons for a lack of definitive conclusion must be stated. ANAB does not dictate how the results are to be communicated or the language to be used, leaving it to the laboratory to determine the proper method based on accepted practice.[359]

ANAB has established *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*. Under "Clear Communications," it requires that ethical and professional forensic scientists present accurate and complete data in reports, testimony, publications and oral presentations. In addition, the *Guiding Principles* state that "reports are

[357] International Laboratory Accreditation Cooperation (ILAC). *Modules in a Forensic Science Process*. Section 4.9.

[358] NIST's National Voluntary Laboratory Accreditation Program (NVLAP) accredits testing and calibration laboratories other than forensic laboratories. It assesses laboratories in compliance with ISO/IEC 17025:2005, and the test report requirements of NVLAP mirror those of the international standards. See National Institute of Standards and Technology (NIST), *Handbook (NIST HB) 150* (2006).

[359] "ISO/IEC 17025:2017(en) General Requirements for the Competence of Testing and Calibration Laboratories," 2017, https://www.iso.org/obp/ui/#iso:std:iso-iec:17025:ed-3:v1:en. Sections 7.8.1.2.2 parts b and c and 7.1.1.I, part 6

prepared in which facts, opinions, and interpretations are clearly distinguishable, and which clearly describe limitations on the methods, interpretations, and opinions reported."[360]

The Bureau of Justice Statistics reported in its *Publicly Funded Forensic Crime Laboratories: Quality Assurance Practices, 2014*[361] that of the 409 publicly funded forensic crime laboratories, 88% were accredited by a professional forensic science organization, which was an increase of 18% in 2002. Seventy-three percent of those laboratories accredited in 2014 were accredited by the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB; now merged into ANAB).[362] In addition to publicly funded crime laboratories, as of April 2019, 49 private corporation laboratories in 57 locations were accredited by ANAB.[363]

The White House Subcommittee on Forensic Science[364] and the NCFS[365] both recommend universal accreditation. Widespread accreditation would ensure that the ISO/IEC 17025:2017 standards on report writing are implemented extensively.

The NCFS recommended a comprehensive report and noted that[366]

> Reports should clearly state: the purpose of the examination or testing; the method and materials used; a description or summary of the data or results; any conclusions derived from those data or results; any discordant results or conclusions; the estimated uncertainty and variability; and possible sources of error and limitations in the method, data, and conclusions.

Found and Bird[367] noted that how each FDE words an opinion varies greatly but typically reflects the probability of a single proposition adopted by the FDE considering the observations of the characteristics in the writing. An alternative approach presented by these authors and recommended by this Working Group (Recommendation 2.5) is to consider "at least two competing and mutually exclusive propositions," and to focus on the evaluation of evidence given each proposition. The FDE conducts the evaluation considering the background information given, the assumptions made, and any limitations present in the evidence. The conclusions may then be expressed as the degree of support for one proposition over other propositions.

Proper interpretation of scientific findings occurs within a framework of circumstances, also known as background information. Evaluations of evidence/findings are conditioned by the proposition(s) and by task-relevant non-scientific case information. The case information is necessary to set appropriate and relevant propositions. It also defines the appropriate population

---

[360] See *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel,* GD 3150, (ANSI National Accreditation Board (ANAB), 2018).

[361] *Publicly Funded Forensic Crime Laboratories: Quality Assurance Practices, 2014,* NCJ 250152, (M. R. Durose et al., 2016).

[362] Durose et al. *Publicly Funded Forensic Crime Laboratories: Quality Assurance Practices, 2014*.

[363] Information provided by ANAB on April 1, 2019.

[364] National Science and Technology Council, Subcommittee on Forensic Science, *Strengthening the Forensic Sciences*, National Science and Technology Council (Washington, DC, 2014), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/forensic_science___may_2014.pdf.

[365] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: Universal Accreditation*, Department of Justice (2015), https://www.justice.gov/archives/ncfs/file/477851/download.

[366] National Commission on Forensic Science (NCFS), *Views of the Commission: Documentation, Case Record and Report Contents*, 2.

[367] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 60.

under the alternative proposition(s) and provides pertinent and relevant information needed (or that is at least beneficial) for a complete evaluation.[368]

Background information is provisional in nature so if the framework information changes, the FDE must reevaluate the findings and adjust their opinion accordingly. For example, if the FDE is told that new information indicates a different underlying writing surface on which the document was written, the FDE may want to reassess their analysis to determine whether the conclusion is still correct based on the new task-relevant information.[369] In general, non-scientific information does not have a direct bearing on the findings; however, it has the potential to bias or influence the interpretation of those findings. This information may be beneficial when it is relevant but is problematic when it is task-irrelevant (see section 2.1). Thus, it is essential to recognize and distinguish between task-relevant versus task-irrelevant information. For example, it may be beneficial to know any unusual conditions relating to the writing act like location, position of the suspect while writing, or unusual activities occurring while writing.

A lack of sufficient task-relevant information may result in poorly formed propositions or the inability to formulate any propositions at all. The report should reflect the propositions used in the evaluation of the evidence and the information that was used to produce them.[370] In addition, the report should indicate that if those propositions change, the opinion of the FDE may also change (see section 2.3.2).

Assumptions are often made by FDEs in terms of the framework information and the nature of the submitted materials. For example, when an FDE uses reference samples to inform their assessment, there are often implicit assumptions about the source of that material or the adequacy and representativeness of the samples.

For example, FDEs may determine that a sample of writing is (1) natural, (2) representative of a writer's habits, and (3) adequate for comparison purposes. It is important to note that this is not an uninformed or naive decision; rather it is "tested" by the FDE in the course of the examination. However, such testing cannot be definitive, and the result is a form of assumption on which the opinion rests, at least in part. Such assumptions have always been made but were generally considered implicit to the process and not expressly stated or acknowledged.

Another common assumption relates to applicability of FDE knowledge to the question at hand. Some FDEs assume their knowledge base is appropriate and adequate for all manner of casework when it is actually best suited to writings with which they are most familiar.

Other assumptions may include that (1) an accurate photocopy or image of the writing (questioned or known) has been provided, (2) the known writing was prepared by the person identified as the writer, or (3) the date of the writing is as purported. It can be difficult to identify

[368] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 60.

[369] ENFSI (2015) notes "Examples of relevant information that could change include the nature of the alleged activities, time interval between incident and the collection of traces (and reference items) and the suspect's/victim's account of their activities." Whether the suspect's/victim's account is task relevant for the analyst depends on the nature of the case and the type of examination being conducted. European Network of Forensic Science Institutes (ENFSI), *Guideline for Evaluative Reporting in Forensic Science*, 59.

[370] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 59.

some types of assumptions; however, when they have been made, such assumptions should be declared to ensure the recipient of the report understands the limits of the opinion.

All of the above points require acknowledgement of the effect of changing framework information. A formal evaluation is conditioned by the propositions and framework information. Because those elements are provisional in nature, it follows that the outcome may change if any of those assumptions change. Similarly, if any of the assumptions made by the FDE are inaccurate, then the evaluation may be affected.

To address this issue, a disclaimer should be provided, such as the following:

> It is important to note that opinions expressed in a report are based upon task-relevant background information and exhibit materials provided to the FDE, as well as the specific propositions used in the evaluation. Should any of the information, exhibit materials, or propositions change, the opinion may also change. In particular, if different propositions are of interest, the FDE should be contacted to discuss the matter further.

The report, then, should state the propositions considered; the background information; and the assumptions, limitations, and conclusions of the examination. Some reports may begin with an executive summary stating the conclusions regarding each document submitted for examination. Other reports are structured in a way that an executive summary is unnecessary.

Although not a part of the report itself, a curriculum vitae (CV) should accompany the report for an analysis of the education, training, experience, and competency of the expert. The CV is also important to determine whether those attributes are relevant to the analysis about which the expert is prepared to testify.

In 2013, Siegal and colleagues surveyed 421 forensic science laboratory reports from 38 publicly funded crime laboratories (in which the directors were members of ASCLD).[371] The report contents were compared with a compilation of report recommendations from 10 forensic science organizations and scientific working groups. The compilation of recommended report contents based on the collected laboratory reports is as follows:

- Demographics: Submitting agency, client, case numbers, charges
- Request for examination: What types of tests are being requested on what evidence
- Inventory of evidence: A listing of what evidence is being submitted
- Executive summary: Akin to a certificate of analysis; what the final conclusions are concerning each piece of evidence submitted
- Methods/materials: A listing of the major chemicals, materials, and instruments used; a listing of the methods used in the analysis of the evidence
- Procedures: Detailed, step-by-step procedures for the analysis of each piece of evidence

[371] Jay A. Siegel, Megan King, and Whitney Reed, "The Laboratory Report Project," *Forensic Science Policy & Management: An International Journal* 4, no. 3-4 (2013), https://doi.org/10.1080/19409044.2013.858798.

- Results: The results of each test run on each piece of submitted evidence
- Discussion: The conclusions reached on the basis of the analysis of each piece of evidence and how each test contributed to the overall conclusions
- Limitations/sources of error: Discussion of the limitations of each test including interfering substances, probative value of the test, specificity, and known sources and rates of errors
- Data: Any charts, graphs, spectra, chromatograms, diagrams, and other data generated by the examination of the evidence
- References: Citations to external written materials used in interpreting the evidence

The project concluded that the reports examined vary widely, based in large part on the type of evidence analyzed and whether the laboratory was federal, state, or local.[372] Many of the reports reflected the testimony before the NRC Forensic Science Committee that "reports are too often more in the nature of certificates of analysis with a short description of the evidence and the results of the analysis, and much less frequently were they true, complete scientific laboratory reports."[373]

With regard to questioned document reports, the project's authors reported that[374]

> Little in the way of methods and procedures is found in these reports. Compared to other types of reports, there is moderate discussion [sic] and limitations/errors. It is somewhat surprising that there is so little in the way of methods and procedures since questioned documents are often subjected to a variety of complex tests.

The criteria against which the 421 laboratory reports were compared were based on ASTM standards and are similar to current ISO/IEC 17025 provisions and accreditation supplemental requirements. The project's conclusions, particularly with respect to questioned document reports, illustrate that there is much room for improvement.[375]

Building on these ideas, the Working Group recommends the following.

> **Recommendation 3.2: At a minimum, the forensic document examiner must include all the information listed below in the case record. Written reports must accurately and clearly detail all relevant aspects of analyses and comparisons. Unless this information is readily accessible by another mode (e.g., case record or report appendices), the written report should include**

[372] Siegel, King, and Reed, "The Laboratory Report Project."

[373] Siegel, King, and Reed, "The Laboratory Report Project," 68. See also pages 71–72.

[374] Siegel, King, and Reed, "The Laboratory Report Project," 74–75.

[375] See "Figures & data" link to review data specific to forensic document examination: https://www.tandfonline.com/doi/figure/10.1080/19409044.2013.858798?scroll=top&needAccess=true (Siegel, King, and Reed, "The Laboratory Report Project.")

1. **Demographics: Submitter, forensic document examiner(s), laboratory, case identifier(s), or other information dictated by the laboratory.**
2. **Request for examination: What is being requested for each document.**
3. **Inventory of evidence: A listing or description of what documents are being submitted, their condition, and unambiguous identification of the items.**
4. **The curriculum vitae for each forensic document examiner.**
5. **A statement of case-related background information provided to the forensic document examiner(s).**
6. **A statement of propositions used in the evaluation of the evidence and a statement that if there are changes to the propositions, the opinion may change.**
7. **A statement of any assumptions made by the forensic document examiner and the basis for them and a statement that if there are changes in the assumptions, the opinion may change.**
8. **Methods: A listing of the instruments and methods used in the examination of the evidence, the range of possible conclusions, and a definition of terms.**
9. **Procedures: Specific step-by-step procedures for the examination of each document or set of documents and any deviations from established test methods.**
10. **Observations: A description of observed characteristics of each document or each set of documents and other bench notes.**
11. **Evaluations: The interpretation of the combined observations given each proposition.**
12. **Conclusions: A complete statement of the conclusions reached based on the observations and evaluations. When associations are made, the significance of the association should be communicated clearly and qualified properly. When exclusions are made, they shall be clearly communicated. When no conclusions are made, the reasons must be clearly stated.**
13. **Limitations: A statement of the limitations of the examination and the procedures.**
14. **Error rates: A statement of potential sources of error and, if available, relevant rates of error; if no relevant error rate is known by the laboratory, that fact should be disclosed.**
15. **Data: Charts, graphs, diagrams, or other data generated by the examination of the evidence as necessary for the proper understanding of the report.**

16. **Review of conclusions: If a review of conclusions occurred, whether a disagreement existed between the forensic document examiner and the reviewer.**

17. **Other statements required by the accreditation body or the laboratory.**

See appendix 3A for a sample report.

3.5. FDE Testimony

The FDE who conducted the examination and wrote the report is the best person to explain the analytical methods and opinions contained in the laboratory report. They may be the only person with the situational awareness of the exact conditions under which the examination was conducted (e.g., mental state of the FDE, working conditions, and cognitive biases that may have affected the conclusion). This is particularly true for handwriting examinations, for which the examination process and conclusions reached have subjective elements.

The testifying expert's personal knowledge of the analysis and the report is important to the education of the fact finder. Such knowledge is also important to the constitutional rights of defendants in criminal cases, as described in *Melendez-Diaz*[376] where the prosecution introduced a laboratory report without the support of a testifying expert. The Supreme Court ruled that the defendant's constitutional right of confrontation was violated. This is not to say, however, that there are no other legitimate methods for presenting forensic evidence when the original reporting expert is unavailable to testify. The evidence can be reanalyzed in some cases, a stipulation can be obtained from the opposing party, or an expert may be able to review the report and case record and arrive at their own opinion. ANAB standards now require that "[t]echnical records to support a test report (including results, opinions, and interpretations) shall be such that, another reviewer possessing the relevant knowledge, skills, and abilities could evaluate what was done and interpret the data."[377] Some states have notice-and-demand statutes that permit the introduction of a certificate of analysis without the presence of the FDE in the absence of the defendant's objection.[378]

The testimony of the reporting expert is also important to litigants in civil cases, because cross-examination in the search for truth is an important element of any litigation involving scientific evidence.[379] Edmond et al. asserted that

> Factors relating to experimental validation, measures of reliability and proficiency are key [elements of cross-examination] because they, rather than conventional legal admissibility heuristics (e.g., field, qualifications, experience, common knowledge, previous admission, etc.), provide information about actual ability and accuracy that enable expert evidence to be rationally evaluated by judges and jurors.[380]

---

[376] Melendez-Diaz v. Massachusetts, U.S.

[377] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Section 7.5.1.3

[378] Melendez-Diaz v. Massachusetts, U.S.. See also *Williams v. Illinois*, 567 U.S. 50 (2012).

[379] *Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.*

[380] G. Edmond et al., "How to Cross-Examine Forensic Scientists: A Guide for Lawyers," *Australian Bar Review* 39 (2014): 174-75. See also K. A. Martire and I. Watkins, "Perception Problems of the Verbal Scale: A Reanalysis and Application of a Membership Function Approach," *Science & Justice: Journal of the Forensic Science Society* 55, no. 4 (Jul 2015), https://doi.org/10.1016/j.scijus.2015.01.002.

In fact, as mentioned earlier, the cross-examination of the expert can be perceived as a form of exploring reliability, or as the NCFS subcommittee has said, a form of "peer review" of the science and the analysis at hand in the legal proceeding.[381] The Supreme Court has agreed, noting that confrontation (cross-examination) is one means of ensuring accurate forensic analysis.[382] If the Supreme Court is correct, then crime laboratories and FDEs should welcome cross-examination because it gives them important feedback on their methods, protocols, and standards.

In *Melendez-Diaz v. Massachusetts*,[383] Justice Scalia suggested four reasons why cross-examination of the expert is important:

1. "Forensic evidence is not uniquely immune from the risk of manipulation. According to a recent study conducted under the auspices of the National Academy of Sciences, '[t]he majority of [laboratories producing forensic evidence] are administered by law enforcement agencies, such as police departments, where the laboratory administrator reports to the head of the agency.'[384] And '[b]ecause forensic scientists often are driven in their work by a need to answer a particular question related to the issues of a particular case, they sometimes face pressure to sacrifice appropriate methodology for the sake of expediency.'[385] A forensic analyst responding to a request from a law enforcement official may feel pressure—or have an incentive—to alter the evidence in a manner favorable to the prosecution."[386]

2. "While it is true . . . that an honest analyst [examiner] will not alter his testimony when forced to confront the defendant [cross-examiner] the same cannot be said of the fraudulent analyst. Like the eyewitness who has fabricated his account to the police, the analyst who provides false results may, under oath in open court, reconsider his false testimony. And, of course, the prospect of confrontation [and cross-examination] will deter fraudulent analysis in the first place."[387]

3. "Confrontation [cross-examination] is designed to weed out not only the fraudulent analyst [examiner], but the incompetent one as well. Serious deficiencies have been found in the forensic evidence used in criminal trials."[388]

4. "Like expert witnesses generally, an analyst's [examiner's] lack of proper training or deficiency in judgment may be disclosed in cross-examination."[389]

---

[381] National Commission on Forensic Science (NCFS), *Views of the Commission: Pretrial Discovery of Forensic Materials*.

[382] Melendez-Diaz v. Massachusetts, U.S.

[383] Melendez-Diaz v. Massachusetts, U.S.

[384] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 183.

[385] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 23–24.

[386] Melendez-Diaz v. Massachusetts, U.S.

[387] Melendez-Diaz v. Massachusetts, U.S.

[388] Melendez-Diaz v. Massachusetts, U.S.

[389] Melendez-Diaz v. Massachusetts, U.S.

In addition, the courts have been designated as gatekeepers regarding expert testimony. To perform that obligation responsibly, the court must carefully examine the contents of the expert's report and their supporting testimony given in a pretrial admissibility hearing. As noted in section 3.6, the courts often use their assessment of the expert's knowledge of the discipline as a critical fact in determining admissibility.

Given those observations, it is the best practice for FDEs who conduct the examination and write the report to testify, when possible. If illness, death, or logistical issues prevent the original FDE from testifying, it is preferable to have the evidence re-examined by a separate FDE who would arrive at their own opinion. The Working Group acknowledges that when either a full review of the case record is conducted or a re-examination is undertaken, the FDE should reduce their cognitive bias by not reviewing the conclusion of the initial FDE before arriving at an independent conclusion. The Working Group recommends the following:

Recommendation 3.3: The forensic document examiner who conducts the examination and writes the report should be the one to testify in any proceeding.

3.5.1. Impartial Testimony

FDEs must testify in a nonpartisan manner and answer questions from all counsel and the court directly, accurately, and fully. They must provide appropriate information before, during, and after trial. That these requirements are necessary for FDEs, and indeed, all forensic scientists, is beyond dispute, and they have accordingly been well established in guiding literature.[390]

The requirement that FDEs be impartial, both as a general matter and in terms of testimony, is appropriately widespread. The ANAB *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel* states that ethical and professionally responsible forensic science personnel and laboratory management "[a]re independent, impartial, detached, and objective, approaching all examinations with due diligence and an open mind."[391] Likewise, to address a recommendation by the NCFS,[392] the Attorney General adopted a Code of Professional Responsibility for the Practice of Forensic Science, which requires forensic practitioners to "[e]nsure interpretations, opinions, and conclusions are supported by sufficient data and minimize influences and biases for or against any party."[393]

The major FDE professional societies expect impartiality from their members in practice and during testimony. The American Society of Questioned Document Examiners (ASQDE) Code of

---

[390] See ANSI National Accreditation Board (ANAB). *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*; "Code of Ethics," n.d., accessed May 6, 2020, https://www.asqde.org/about/code_of_ethics.html; American Board of Forensic Document Examiners (ABFDE), *Code of Ethics and Standard Practices* (2014), https://www.abfde.org/htdocs/AboutABFDE/Ethics.pdf; Scientific Working Group on Friction Ridge Analysis, Study and Technology, *A Model Policy for Friction Ridge Examiner Professional Conduct*, Scientific Working Group on Friction Ridge Analysis, Study and Technology (2008); The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 117.

[391] See ANSI National Accreditation Board (ANAB). *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*, 1.

[392] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: National Code of Professional Responsibility for Forensic Science and Forensic Medicine Service Providers*, Department of Justice (2016), https://www.justice.gov/ncfs/file/839711/download.

[393] Attorney General, *Memorandum for Heads of Department Components: Recommendation of the National Commission on Forensic Science; Announcement for NSFS Meeting Eleven* (2016), https://www.justice.gov/opa/file/891366/download.

Ethics states that its members must agree "to act at all times, both in and out of court in an absolutely impartial manner and to do nothing that would imply partisanship or any interest in the case except to report the findings of an examination and their proper interpretation."[394] The Association of Forensic Document Examiners (AFDE) Code of Ethics also requires its members to base their findings and opinions in every case "solely upon the facts and merits of the evidence [they] have examined," to "seek to understand the truth, without bias, for or against any party," and to "communicate [their] findings and opinions as clearly and fairly as [they are] able."[395] Both professional associations have procedures in place to address complaints, allegations, or charges like oral or written reprimand, suspension, or termination.

The BFDE Code of Ethics and Professional Responsibility also requires that its Diplomates "render opinions that are clearly supported by the evidence examined" and "[undertake] each assignment objectively and solely with a view towards ascertaining demonstrable facts from which an opinion may properly be derived, without bias as to the outcome."[396] The Code of Ethics and Standard Practices for the ABFDE (ABFDE Code) likewise requires that "[a] Diplomate or candidate of the ABFDE will only render opinions . . . which are within his/her area of expertise, and will act, at all times, in a completely impartial manner by employing scientific methodology to reach logical and unbiased conclusions."[397] The Working Group notes that although the scientific method can (and typically does) promote impartiality, its use does not guarantee that testimony will be given in an impartial manner; even results found through valid scientific means may be unfairly communicated to a fact finder. Thus, the Working Group suggests that requirements for impartiality in testimony and the use of the scientific method be made explicit in any code of conduct.

Distinct from, but related to, impartiality is the requirement that all testimony, like the examination and conclusion to which it pertains, "[e]nsure interpretations, opinions, and conclusions are supported by sufficient data."[398] An expert should, moreover, "clearly distinguish data from interpretations, opinions, and conclusions."[399] This provision helps different components of testimony be properly understood and weighed. Also key in this regard is the expert's discussion of uncertainty. Like all forensic disciplines, forensic handwriting examination has sources of error, uncertainty, and limitations.[400] Therefore, testimony should include discussions of these topics.

To that end, the NRC report recommended that expert testimony include "as appropriate, the sources of uncertainty in the procedures and conclusions along with estimates of their

---

[394] American Society of Questioned Document Examiners (ASQDE), "Code of Ethics." Item (e)

[395] "Code of Ethics," n.d., accessed May 6, 2020, https://afde.org/resources/AFDE_CODE-OF-ETHICS.pdf.

[396] American Board of Forensic Document Examiners (ABFDE), *Code of Ethics and Standard Practices*. Paragraphs 3.1.3 and 4.1.1. See also paragraph 5.1, Integrity Related to Examination Procedures, and paragraph 5.2, Integrity Related to Opinion and Conclusions.

[397] American Board of Forensic Document Examiners (ABFDE), *Code of Ethics and Standard Practices*. Rule 8

[398] U.S. Department of Justice, *Code of Professional Responsibility for the Practice of Forensic Science* (2016), https://www.justice.gov/sites/default/files/code_of_professional_responsibility_for-the_practice_of_forensic_science_08242016.pdf. See also National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: National Code of Professional Responsibility for Forensic Science and Forensic Medicine Service Providers*. See paragraph 5 (experts should "[u]tilize scientifically validated methods and new technologies, while guarding against the use of unproven methods in casework and the misapplication of generally-accepted standards").

[399] U.S. Department of Justice, *Code of Professional Responsibility for the Practice of Forensic Science*. Paragraph 12.

[400] See Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83. ("There are limitations associated with the comparison of handwriting for use in forensic science." p. 9).

[significance] (to indicate the level of confidence in the results)."[401] The Department of Justice *Code of Professional Responsibility for the Practice of Forensic Science* also recommends that practitioners disclose "known limitations that are necessary to understand the significance of the findings."[402] Similarly, the ASQDE Code states that members must "render an opinion or conclusion strictly in accordance with the physical evidence in the document, and only to the extent justified by the facts" and "[t]o admit frankly that certain questions cannot be answered because of the nature of the problem, the lack [of] material, or insufficient opportunity for examination."[403] The BFDE requires that its certificate holders "[a]ccurately and honestly report […] all results or data obtained from examining evidence."[404] These rules, properly understood and applied, should lead to appropriate testimony and should include the level of empirical support that exists for any method described in the report.

Reporting this information is necessary to ensure that testimony is appropriately understood and properly weighed. To the extent that the error rate or the significance of uncertainty is unknown, those facts must also be reported to the fact finder in both reporting and testimony. The Working Group suggests that estimates of error rate be developed so that FDEs are able to provide them during testimony.[405] Impartial testimony, supported by science, implicitly requires an FDE to answer questions from all counsel and the court directly, accurately, and fully. In an adversarial system, the parties have distinct ethical obligations and roles, which may incentivize them to ask questions and seek testimony that benefits their side,[406] and, in fact, under this system FDEs are called "for" a particular side. But despite the pressures inherent in such a system, the FDE's overriding duty regardless of which side calls them or of any attempts by counsel (or even the court) to misconstrue or overstate testimony, is to remain impartial and to "[p]resent accurate and complete data in reports, testimony, publications and oral presentations."[407]

For example, if FDEs are required to answer yes or no to a question, they should "[a]ttempt to qualify their responses while testifying" if failing to do so "would be misleading to the judge or the jury."[408] The BFDE counsels the same in its Code, stating that FDEs shall "reject any suggestion, pressure or coercion to render an opinion that is misleading or inconsistent with the examiner's findings"[409] and "[i]f an opinion requires or warrants qualification or explanation so that the opinion is not overstated, misconstrued, or misunderstood, it is not only proper for, but

---

[401] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 21.

[402] U.S. Department of Justice, *Code of Professional Responsibility for the Practice of Forensic Science*. Paragraph 12.

[403] American Society of Questioned Document Examiners (ASQDE), "Code of Ethics." Item (e).

[404] "Ethics," 2012, accessed May 6, 2020, https://www.bfde.org/ethics.html. Paragraph 4.1.3.

[405] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 5–6.

[406] Lawyers, for example, owe a duty to their clients to "act with commitment and dedication to the interests of the client and with zeal in advocacy upon the client's behalf." See "Rule 1.3 Diligence—Comment," n.d., accessed May 7, 2020, https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_3_diligence/comment_on_rule_1_3.html. Criminal defense lawyers and public prosecutors also have special duties and responsibilities that may sometimes put them at odds with a forensic practitioner. *E.g., id.* at Rule 3.1 (noting that while lawyers may not bring frivolous claims, "[a] lawyer for the defendant in a criminal proceeding, or the respondent in a proceeding that could result in incarceration, may nevertheless so defend the proceeding as to require that every element of the case be established."); *id.* at Rule 3.8 (describing special duties of prosecutors).

[407] See ANSI National Accreditation Board (ANAB). *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*. Paragraph 14.

[408] ANSI National Accreditation Board (ANAB). *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*. Paragraph 19.

[409] Board of Forensic Document Examiners (BFDE), "Ethics." Paragraph 5.2.1.1.

also is *incumbent* upon, the forensic document examiner to offer such qualification" (emphasis added).[410]

For its part, the ENFSI expects FDEs to ensure that they "[d]eal with questions truthfully, impartially and flexibly in a language which is concise, unambiguous, and admissible."[411] All FDEs should therefore use, as the NRC report[412] advises, plain language so that all trial participants are able to understand and appropriately weigh the testimony. Such "clear and straightforward terminology"[413] may help promote the appropriate use and understanding of handwriting examination by other stakeholders in the system. However, the Working Group acknowledges that it is not easy to determine terminology that is "clear and straightforward," and that more research is needed to assess how terminology used by the FDE is interpreted by the fact finder. Finally, the FDE should "[h]onestly communicate with all parties (the investigator, prosecutor, defense, and other expert witnesses) about all information relating to his or her analyses, when communications are permitted by law and agency practice."[414]

Human factor issues relating to communication beyond testimony are discussed in chapter 4, box 4.1 (Duty to Correct) and section 6.3.3 (Communication with Other Stakeholders).

Recommendation 3.4: Forensic document examiners must testify in a nonpartisan manner; answer questions from all counsel and the court directly, accurately, and fully; and provide appropriate information before, during, and after trial. All opinions must include an explanation of any data or information relied on to form the opinion.

3.5.2. Reporting the Possibility of Error

Although the use of a robust QA system should reduce the magnitude and frequency of errors (see section 4.2 for more information on QA systems), it is an FDE's duty to acknowledge, in both written and oral reports and testimony, that the possibility of error exists.

According to Budowle et al.,[415]

> An examiner may not state or imply that the method used has a zero error rate or is infallible, due to the possibility of practitioner error. A testifying expert should be prepared to describe the steps taken in the examination process to reduce the risk of observational and judgmental error. However, the expert should not state that examiner errors are inherently impossible or that a method inherently has a zero error rate. The literature related to error rates emphasizes

[410] Board of Forensic Document Examiners (BFDE), "Ethics." Paragraph 5.3.1.3.1.

[411] *Performance Based Standards for Forensic Science Practitioners,* (European Network of Forensic Science Institutes (ENFSI) Standing Committee for Quality and Competence, 2004), 43. Standard I3 (d).

[412] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 186. The NAS Report further underscores the need for more substantial research in this regard so that the reliability of different methods and their associated confidence intervals can be understood.

[413] U.S. Department of Justice, *Code of Professional Responsibility for the Practice of Forensic Science*. Paragraph 12 (recommending that forensic practitioners "[p]repare reports and testify using clear and straightforward terminology").

[414] ANSI National Accreditation Board (ANAB). *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*. Paragraph 4.

[415] B. Budowle et al., "A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement," *Journal of Forensic Sciences* 54, no. 4 (Jul 2009), https://doi.org/10.1111/j.1556-4029.2009.01081.x.

the difficulty in calculating a meaningful error rate for both individual practitioners, as well as across the entire discipline.

Because the possibility for practitioner error exists, it is important for an FDE to understand and demonstrate to the fact finder how QA measures help reduce the risk of error in the examination process. Verification of an FDE's conclusions is one of those important quality measures. However, one state appellate court has ruled that testimony before the jury concerning verification in the particular case by a non-testifying expert is inappropriate bolstering of the testifying expert.[416] Testimony before the jury about verification in the case has to be carefully crafted to avoid allegations of bolstering. Of course, such testimony would be unobjectionable in a *Daubert*[417] hearing because verification speaks to reliability—one of the determinations to be made in such a hearing—and because the rules of evidence[418] do not apply.

Regarding the determination of error rates for forensic handwriting examination, Found and Bird[419] posited that although some individuals may try to derive a global error rate for forensic handwriting examination about all types of writing and all FDEs in general, this is not an appropriate position to take. This rationale is derived from two main sources. First,

> all validation studies to date have shown that examiners [sic] responses on blind trials vary, and can vary widely, particularly in terms of individuals' correct and inconclusive scores. Therefore the results from one group of examiners or an individual examiner may not be a good estimate of the potential results of an unrelated group or individual in spite of these examiners using the same resource materials, being the product of similar training regimes and even using similar methodology (see section 2.2.2). As a human skill this is not entirely unexpected.[420]

Second,

> in the majority of instances, questioned writing can be either normal writing by the specimen writer, disguised writing by the specimen writer, auto-simulated writing, normal writing not by the specimen writer, disguised writing not by the specimen writer or simulated writing not by the specimen writer (forgeries). . . . Since there are a number of different categories of questioned writing, there is the real possibility that the potential error for opinions expressed within each of these categories may be different.[421]

Research by Found and Rogers[422] suggests a global estimate of error would be a skewed one, based on the numbers of each writing category. As such, "this is problematic and must be taken

---

[416] *Miller v. State*, 127 So.3d 580 (Fla. Dist. Ct. App. 2012).

[417] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.

[418] Federal Rules of Evidence 104(a).

[419] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 64.

[420] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version."

[421] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 64.

[422] B. Found and D. Rogers, "Problem Types of Questioned Handwritten Text for Forensic Document Examiners" (International Graphonomics Society Biennial Conference, Salerno, Italy, Editrice Zona, Arezzo, Italy, June 26–29 2005); B. Found and D. Rogers, "The Probative Character of Forensic Handwriting Examiners' Identification and Elimination Opinions on Questioned Signatures," *Forensic Science International* 178, no. 1 (Jun 10 2008), https://doi.org/10.1016/j.forsciint.2008.02.001.

into consideration when arriving at a philosophy of potential error estimation."[423] It may be possible to partially mitigate this issue if the FDE addresses each of the relevant propositions (or sub-propositions) with error estimates generally relating to the different types of writing or writing conditions. It may then be possible to delineate different error estimates and apply them to the assessment process. See section 4.2.6.7 to 4.2.6.9, for discussion on delineating different error estimates.

The FDE should be prepared during testimony to describe any steps taken during the examination process that lessened the potential for biasing effects to influence the opinion regarding the examined evidence. These steps include the adoption of CIM into procedures. This is thoroughly discussed in section 2.1. To summarize, the FDE should have minimal exposure to task-irrelevant information in a case and should be transparent in both the report and testimony when they are exposed to such information.

Recommendation 3.5: In testimony, a forensic document examiner must be prepared to describe the steps taken during the examination to reduce the risk of process, observational, and cognitive errors. The forensic document examiner must not state that errors are impossible.

3.6. The FDE's Knowledge of the Discipline

FDEs have the responsibility to support the admissibility of handwriting examination when answering questions from an attorney or judge. Knowledge of underlying principles and research enables the expert to answer questions regarding the *Daubert*[424] factors and requirements of FRE 702. A working knowledge of the relevant research should include the ability to describe the sample size of any referenced studies and the composition, study test conditions, and specific findings. This information can be helpful to the court in determining any "analytical gap between the data [in the studies] and the opinion offered"[425] is not unreasonable. If the expert cannot address such questions, the judge may lack sufficient supportive information on which to rule in favor of admissibility.

Indeed, there have been cases in which an expert's insufficient knowledge of the underlying principles and research may have contributed to rulings against admissibility. For example, in *United States v. Saelee*,[426] the court noted that

> [the expert] testified that he did not know whether any of the articles discussed error rates, empirical testing, or coincidental matches, although he claimed to have read the articles. The list, without analysis of the substance of the articles, is of little use to the court.

In *United States v. Lewis*,[427] the court observed that the "[expert] could not testify about the substance of the studies he cited. He did not know the relevant methodologies or the error rates

[423] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83.

[424] Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S.

[425] General Electric Co. v. Joiner, U.S.

[426] *United States v. Saelee*, 162 F. Supp. 2d 1097, 1103 (D. Alaska 2001).

[427] United States v. Lewis, F. Supp. 2d.

involved in these studies."[428] Accordingly, the court concluded that the expert's "bald assertion that the 'basic principle of handwriting identification has been proven time and time again through research in [his] field,' without more specific substance, is inadequate to demonstrate testability and error rate."[429]

Likewise, in *United States v. Johnsted*,[430] the court concluded that "the government ha[d] not provided enough evidence to demonstrate the reliability of handwriting analysis to the hand printing in this case." In so finding, the court wrote that

> The government's decision to provide nothing more than [the expert's] single-sentence conclusion, and in particular to provide no explanation of the underlying basis for her conclusion, leaves the court with nothing to hang its hat on in determining whether [the expert's] methodology and analysis in this case are supported by scientifically valid principles.[431]

More research is needed about the assumptions and principles underlying the elements of forensic handwriting examinations, and FDEs will need to continually update their familiarity with new research (see section 2.3.3).

> **Recommendation 3.6: Forensic document examiners must have a functional knowledge of the underlying scientific principles and research regarding handwriting examination, as well as reported error rates or other measures of performance, and be prepared to describe these during their testimony.**

3.7. Use of Visual Aids During Testimony

Human beings are visually oriented creatures, and much of the information about the world around us comes in the form of visual input. In general, humans are adept at pattern-matching and similar recognition tasks. When addressing visual evidentiary material (or material that is latent but able to be visualized), it follows that demonstrative aids can be very helpful when explaining the basis for an opinion. Indeed, studies have shown that visual aids may increase understanding and retention levels of oral testimony by up to 65%.[432] Visual evidence "is generally more effective than a description given by a witness, for it enables the jury, or the court, to see and thereby better understand the question or issue involved."[433] Enhancing the fact finders' understanding of the evidence is important because "crucial evidence can be rendered useless or even a liability if the jury does not understand the evidence or appreciate its significance."[434]

---

[428] United States v. Lewis, F. Supp. 2d.

[429] See also *United States v. Lewis, F. Supp. 2d.*. "[Expert] had no explanation for why twenty-five samples of writing were necessary for a comparison of handwriting. He simply said that twenty-five samples was the number generally used."

[430] *United States v. Johnsted*, 30 F. Supp. 3d 814, 821 (W.D. Wis. 2013).

[431] United States v. Johnsted, F. Supp. 3d at 821.

[432] Karen D. Butera, "Seeing is Believing: A Practitioner's Guide to the Admissibility of Demonstrative Computer Evidence," *Cleveland State Law Review* 46, no. 3 (1998).

[433] *Alston v. Shiver*, 105 So. 785 (Flo. 1958).

[434] M. Q. Cooper, "Practitioner's Guide: the Use of Demonstrative Exhibits at Trial," *Tulsa Law Journal* 34, no. 3 (1999).

Visual material can help the viewer understand the information being presented. It should be designed so that the viewer can (1) see the features of interest, (2) better understand the features of interest, and (3) more fully appreciate subtleties in the features that would otherwise be obscured.

Handwriting is a dynamic physical action that produces a static, visual record familiar to most people. Familiarity with handwriting by laymen is both a blessing and a curse to the FDE and the legal system. On one hand, because people are familiar with handwriting, they can readily understand the FDE's explanation if it is given clearly and in terms that make sense to them. On the other hand, people might presume that they understand more than they do even though they are not educated in the principles that underlie handwriting examination unless informed by the FDE through testimony.

Visual demonstrations prepared by the FDE help educate the jury. Graham wrote that "[d]emonstrative evidence... is distinguished from real evidence in that it has no probative value in itself, but serves merely as a visual aid to the jury in comprehending the verbal testimony of a witness."[435] This definition of demonstrative evidence is consistent with the court's use of the term in *Baugh ex rel. Baugh v. Cuprum S.A.de C.V.*,[436] which, recognizing the ambiguity in the term and its various uses in the courts, defined "'demonstrative' [to signify] that the exhibit is not itself evidence—the exhibit is instead a persuasive, pedagogical tool created and used by a party as part of the adversarial process to persuade the jury."[437]

Demonstrative evidence may include pedagogical charts or summaries of a witness's conclusions or opinions, "or they may reveal inferences drawn in a way that would assist the jury," but "displaying such charts is always under the supervision of the district court under Rule 611(a), and in the end are not admitted as evidence."[438] FRE 611(a) gives a judge discretion over the use of demonstrative evidence in controlling the mode and order of presenting evidence, including whether the presentation of demonstrative evidence is "effective for determining the truth."[439]

A court has the duty to determine whether the demonstrative evidence accurately reflects the evidence presented. Demonstrative aids, whether incorporated into work notes, the report, or produced solely for court presentation purposes must be prepared in a manner that accurately represents the information. In particular, the aids should be consistent with the report and present a fair, objective, and unbiased view of the evidence. The demonstrative exhibits must be focused on elements relevant to the testimony of the expert and consistent with the expert's report and must not be unfairly prejudicial, confusing, or misleading.

Demonstrative aids can be double-edged swords. Although a good visual aid can assist the viewer in understanding a forensic examiner, a poorly prepared aid may confuse the viewer or provide a biased perspective on the matter by taking information out of its original context.

---

[435] Michael H. Graham, "Real and Demonstrative Evidence, Experiments and Views," *Criminal Law Bulletin* 46, no. 4 (2010), https://ssrn.com/abstract=1885714.

[436] *Baugh ex rel. Baugh v. Cuprum S.A.de C.V.*, 730 F.3d 701 (7th Cir. 2013).

[437] Baugh ex rel. Baugh v. Cuprum S.A.de C.V., F.3d.

[438] *United States v. Janati*, 374 F.3d 263, 273 (4th Cir. 2004); Baugh ex rel. Baugh v. Cuprum S.A.de C.V., F.3d.

[439] A comprehensive discussion of demonstrative evidence can be found in M. Howard and J. Barnum, "Bringing Demonstrative Evidence in from the Cold: The Academy's Role in Developing Model Rules," *Temple Law Review* 88, no. 3 (2016), https://ssrn.com/abstract=2710182.

Demonstrative visual aids generally summarize the material being depicted while reorganizing it into some new form or layout.

A careless or biased presentation could result in an exhibit that presents a misleading view. For example, if only carefully selected known signatures are presented with a questioned signature, a judge or juror might be misled into thinking that a particular feature did not appear in the known writing, when in fact it did. Similarly, if single letters are compared in isolation, the placement of the letter within a word or the connection to other letters could be misrepresented. Such features may be important and may be inconsistent with the FDE's conclusion, although unnoticed by the viewer because of the way the aid was presented to them.

A proactive practice would be for the FDE to include images of features that could raise questions about the opinion and explain why the opinion was reached while addressing those questions. In addition, standard procedures—like including a measurement scale and keeping all images in proportion to that scale—are important, particularly if measurements are included in the basis for the opinion.

The Working Group therefore recommends the following.

> **Recommendation 3.7: Demonstrative visual aids, when used, must be consistent with the report and anticipated verbal testimony. Aids must accurately represent the evidence, including both similarities and dissimilarities found in samples and be prepared and presented in a manner that does not misrepresent, bias, or skew the information.**

Appendix 3A: Sample Report

This appendix provides an example of a report that includes all of the information required in Recommendation 3.2. It is not presented as a mandatory structure or layout. Callout boxes reference the information type as outlined in Recommendation 3.2. Note that the report refers to three attachments; however, only the illustration is attached for this example. The report uses a likelihood ratio approach to evidence evaluation and reporting.

Susan Whitford
a. Examiner/laboratory
Phone: 555-555-5555
Forensic Document Examiner
Fax: 888-888-8888
P.O. Box 1234
Email: susan@susanwhitford.com
Boston, MA

SAMPLE REPORT ON THE EXAMINATION OF HANDWRITING

To: Mr. Roger Brown
Brown and Green, PLLC
Boston, MA
a. Submitter
Date: April 21, 2017

Case Number: 17-0018
a. Case identifier

1. Items received
c. Inventory of evidence

The following documents were received from Mr. Robert Brown, Brown and Green, PLLC, on March 27, 2017, and were specified as having known or questioned signatures.

Item #	Type of Document	Date	Known or Questioned
K1	Promissory note, in the amount of $16,500.00	3/18/15	Known signature of Edna Wilson
K2	Insurance application, page 3	3/26/15	Known signature of Edna Wilson
K3	Request for petty cash reimbursement	5/17/15	Known signature of Edna Wilson
K4	Delivery receipt	11/3/15	Known signature of Edna Wilson
K5	Project report, section 7b	1/8/16	Known signature of Edna Wilson
K6	Fax cover sheet, to James River Landscaping	3/30/16	Known signature of Edna Wilson
K7	Fax cover sheet, to ABC Pools	3/30/16	Known signature of Edna Wilson
K8	Interoffice memo, to “Claire Henderson”	4/14/16	Known signature of Edna Wilson
K9	Change of beneficiary form	5/10/16	Known signature of Edna Wilson
K10	Affidavit	5/12/16	Known signature of Edna Wilson
K11	Interoffice memo, to “Claire Henderson”	6/2/16	Known signature of Edna Wilson
Q1	Letter, to Prosecutor David Smith	2/1/16	Questioned signature of Edna Wilson

2. Information obtained e. Statement of background

Attached is the letter of instruction that accompanied the documents for examination from Brown and Green, PLLC.

3. Examination requested b. Request for examination

To determine whether or not Edna Wilson, known signer of documents K1–K11 listed above, signed the questioned document, Q1.

4. Propositions f. Statement of propositions

The following two mutually exclusive propositions were formulated for the questioned signature before examination:

P1. The signature "Edna Wilson" on questioned document Q1 was written by Edna Wilson.

P2. The signature "Edna Wilson" on questioned document Q1 was written by someone other than Edna Wilson.

5. Procedures h. Method and i. Procedures

The original documents were examined with a stereo zoom microscope. The documents were also scanned at a resolution of 600 dpi. The questioned and then the known signatures (and enlargements of these) were examined individually and then compared. Standard document examination methodology was followed.[440] Portions of the documents were extracted and arranged in a chart attached to this report as Illustration 1.

6. Error Rate n. Error rates

Error rate estimates relevant to the examination procedures used have been reported and presented in these peer-reviewed studies [list relevant studies to the examination performed]. In general, testing and evaluation of the examination process done to date on the specific claims addressed in these studies the accuracy has been found to be generally high in settings similar to this case; however, please note that the references to error rates are only presented to verify the general validity and accuracy of the methods used in this examination and do not directly reflect the evidential value of the recovered evidence. Please see section 8 for a summary of the evidential value.

[440] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83.

Observations

j. Observations

a. Questioned material

The questioned letter Q1 contains an original ink signature in the name "Edna Wilson" and is dated February 1, 2016. The signature is a sufficient writing sample to warrant a forensic examination. The signature is what can be described as text-based, with the letters "Edna Wilson" legible. There are three pen lifts within the signature: after "d" and "a" in "Edna" and after the "W" of "Wilson," and there is some tapering of the commencement and terminal strokes and variation in pen pressure, indicating the signature was written with reasonable speed. The signature displays a forehand slope, with the baseline of the signature rising to the right. It has been reproduced at the top left of Illustration 1.

b. Known material

Eleven known signatures of Edna Wilson appear on various original documents written in the course of day-to-day life. These are dated between March 18, 2015, and June 2, 2016, a time period that spans the date of the questioned document. The known signatures can be classified as text-based, with the letters "Edna Wilson" largely legible in each signature. The signatures display a forehand slope, with the signature baseline usually rising to the right (although K2 and K3 have largely a horizontal baseline). Connectivity within the known signatures varies. Typically, the "Ed" "na" "il" and "son" letter combinations are connected. In one of the signatures (K3), the letters "Edn" are connected, in another (K11), the letters "Edna" are connected, and in K3, K4, and K11 all of the letters after "W" are connected. The "s" in "Wilson" varies in formation from a cursive style (K3, K4, and K11) to a more hand-printed style. Taken together, the eleven known signatures provide a reasonable insight into the normal variation in the signatures of Edna Wilson over the period represented. They are reproduced in chronological order in Illustration 1.

7. Results of the comparison

Compared with the known signatures of Edna Wilson during the same time period, similarities were observed in the overall design, proportions, connectivity, and details of construction.

1. General slant to the right of vertical.
2. Text-based (legible) style of the signature.
3. Construction of the "E" of "Edna"—The use of the Greek "E" with the top of the "E" and the terminal stroke of the "E" being diagonally oriented.
4. "Ed" connection—The "E" connects to the "d" of "Edna" at the top of the bowl of the "d."
5. Construction of the "d" in "Edna"—The body of the "d" is thin and diagonally oriented. The stem of the "d" is looped.
6. Pen lifts after "d" of "Edna" and the "W" of "Wilson."

7. Construction of the "W" of "Wilson"—The simple "W" with rounded turning points.

8. Proportions—The height difference between the upper and lower case letters. No significant differences were observed.

8. Interpretation of the findings of the examination

k. Evaluations

The questioned signature appears to have been written with reasonable speed and displays similarities to the known signatures in regard to its overall design, slant, and complexity. Similarities in the finer details of construction, proportions, and connectivity were also observed. This degree of correspondence is what I expect if two pieces of writing were by one person, and therefore, I consider that the probability of these combined findings is high if the questioned signature on Q1 was written by Edna Wilson (P1). In other words, the findings provide very strong support for P1 considered on its own.

From my experience and training, I consider that the combination of features observed is not common, and these observations are not what I expect if the questioned signature was written by someone other than Edna Wilson (P2). Therefore, the probability of observing the degree of similarity given the questioned signature was written by someone other than Edna Wilson is assessed to be low. The findings provide very little support for P2 considered on its own.

The findings, therefore, are much more likely if P1 is true than if P2 is true. In other words, this implies that the findings provide much greater support for P1 than for P2.

9. Conclusion

l. Conclusions

It is my opinion that the evidence observed provides very strong support for the proposition that the questioned signature was written by Edna Wilson over the proposition that the questioned signature was written by someone other than Edna Wilson.

My opinion is based on the information and material submitted to me and is based upon the specific propositions outlined above. Should this information, exhibit material, or the propositions change, my opinion may also change. In particular, if different propositions are of interest, I should be contacted to discuss the matter further.

10. Assumptions

g. Statement of assumptions

I have assumed that the purported dates on each of the known and questioned documents are correct. I have also assumed that the signatures submitted as known writings of Edna Wilson (K1–K11) are indeed writings of that person and that they display the normal variation in the signatures of Edna Wilson over the period represented.

11. Limitations

m. Limitations

In some cases, there are limitations to an examination that require the FDE to state a qualified opinion. Such limitations include insufficient or incomparable known samples, poor quality of questioned or known writing, and lack of complexity in the questioned writing. In the case at hand, there were no such limitations to the examination.

12. Additional information

p. Review of conclusions

The case file associated with this examination, including my conclusions and report, have not been subjected to a technical review.

Susan Whitford

Attachments

1. Letter of instruction from Brown and Green, PLLC

e. Statement of background

2. Illustration 1

3. CV of Susan Whitford

d. CV of examiner

Illustration 1

o. Data

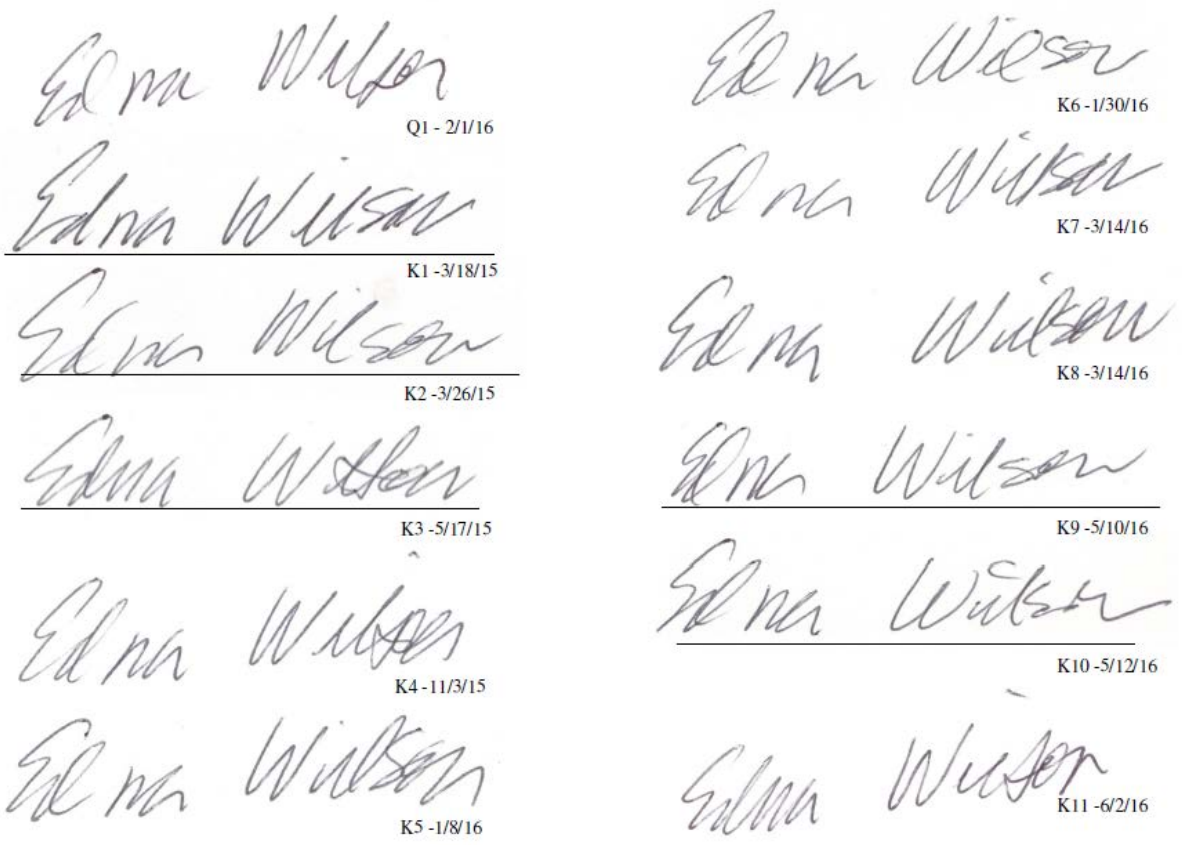


Illustration 1 - The questioned signature is at the top left. The known signatures are arranged in chronological order. The handwriting features are discussed in the Observations section of this report.

Prepared by Susan Whitford, FDE Illustration 1

Appendix 1. Opinion scale

The opinion scale used is detailed in The Modular Forensic Handwriting Method.[441] Conclusions are intended to convey the degree of support provided by the observed evidence for one proposition versus another proposition. The levels available are listed below.

[441] Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 7–83.

1. The evidence provides very strong support for proposition X over proposition Y.
2. The evidence provides qualified support for proposition X over proposition Y.
3. The evidence provided approximately equal support for propositions X and Y.
4. The examination was inconclusive (when limitations in the submitted material severely limit/preclude the examination).

This page intentionally left blank.

4. QA/QC

Introduction and Scope

A QA/QC program organizes, documents, and promotes consistency and accuracy in the work product. Because QA/QC provides the backbone for all efforts to identify, understand, mitigate, and help prevent errors in the forensic sciences. This chapter details the basic requirements to set up and oversee human factors aspects of a QA/QC program.

QA focuses on planning procedures to prevent error whereas QC focuses on monitoring the activities for error detection. QA relies on feedback from QC. In this chapter, the combined efforts of QA and QC are referred to as the Quality Management System (QMS). A laboratory's QMS consists of policies, procedures, and practices, outlined in a quality manual, to evaluate and improve the activities of personnel. The system is most effective when management and employees are devoted to its implementation and continual improvement.

One of the most important tenets of the human factors domain is timely feedback.[442] In the absence of a robust QMS, FDEs may not be given the opportunity to obtain this feedback and thus mitigate potential issues that may later become evident during trial or other inopportune times. Both public and private laboratories stand to benefit from such a program.

Accreditation is intended to be an external check of laboratories to determine if they are performing competent work as outlined in their standard operating procedures and in compliance with accreditation standards.[443] This chapter outlines the requirements and benefits of accreditation and the associated QMS. This chapter also highlights how accreditation and QMS elements can assist in reducing the potential for error in laboratory practices.

[442] See G. Hardavella et al., "How to Give and Receive Feedback Effectively," *Breathe (Sheff)* 13, no. 4 (Dec 2017), https://doi.org/10.1183/20734735.009917.{Hattie, 2016 #152; John Hattie and Helen Timperley, "The Power of Feedback," *Review of Educational Research* 77, no. 1 (2016), https://doi.org/10.3102/003465430298487; G. D. Schiff, "Minimizing Diagnostic Error: the Importance of Follow-Up and Feedback," *The American Journal of Medicine* 121, no. 5 Suppl (May 2008), https://doi.org/10.1016/j.amjmed.2008.02.004.

[443] "About ILAC," 2019, accessed May 6, 2020, https://ilac.org/about-ilac/.

4.1. Accreditation

Crime laboratory accreditation has been one of the most significant developments for American crime laboratories in the last 3 decades.[444] Effective QA programs are the foundation for good forensic science, reliable techniques to apply the science, and trustworthy expert testimony.[445] Encouraged by judicial opinions,[446] mandated by state legislatures,[447] and implemented by crime laboratory directors,[448] accreditation programs have brought needed oversight to a critical segment of our criminal justice system.[449] The use of consensus-based international standards like those produced by the ISO/IEC, in an independent accreditation process, addresses previous criticism that crime laboratory accreditation programs are designed, adopted, implemented, and overseen solely by laboratory personnel. The ISO/IEC guidance and requirements documents are internationally developed and accepted accreditation standards.[450]

Virtually every report that discusses laboratory accreditation as part of a QMS has recognized its importance. The 1992 NRC report suggested that courts should view the absence of appropriate accreditation as constituting a prima facie case that the laboratory has not complied with generally accepted standards.[451] In 1997, the Department of Justice Office of Inspector General report[452] on its investigation of allegations concerning the FBI laboratory recommended that the FBI laboratory obtain accreditation by American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB) as soon as possible. A 2006 report by the American Bar Association Criminal Justice Section recommended that "crime laboratories and medical examiner offices should be accredited, examiners should be certified, and procedures should be standardized and published to ensure the validity, reliability, and timely analysis of forensic evidence."[453]

---

[444] ASCLD/LAB received its first accreditation applications in early 1982. See K. Melson, "Crime Laboratory Accreditation: The Assurance of Quality in the Forensic Sciences," in *The State of Criminal Justice* (American Bar Association, 2003), 3. For a history of accreditation development in the United States, see National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 197.

[445] The elements that make up a comprehensive quality assurance program are described in National Research Council (NRC), *DNA Technology in Forensic Science*, The National Academies (Washington, DC, 1992), 98.

[446] In Daubert v. Merrell Dow Pharmaceuticals, Inc., U.S., the Supreme Court noted that a court ordinarily should consider the existence and maintenance of standards controlling the technique's operation when determining admissibility of scientific evidence (citing *United States v. Williams*, 583 F.2d 1194, 1198 (2nd Cir. 1978).) (noting professional organization's standards governing the technique). Judges are citing the accreditation standards in decisions on admissibility of scientific evidence. See, e.g., *Smith v. State*, 702 N.E.2d 668, 673 (Ind. 1998); Williams v. Illinois, U.S. (noting the use at trial of a DNA report prepared by a modern, accredited laboratory); and *United States v. Anderson*, 169 F.Supp.3d 60 (D.D.C. 2016).

[447] As of 2013, 13 states and the District of Columbia had passed legislation mandating accreditation and other oversight requirements for at least some forensic service providers, including Arkansas, California, Hawaii, Indiana, Louisiana, Maryland, Missouri, Nebraska, New York, North Carolina, Oklahoma, and Texas. https://www.ncsl.org/Documents/cj/AccreditationOfForensicLaboratories.pdf; Accreditation is required only for laboratories conducting forensic DNA analysis in California, Hawaii, Indiana, and Nebraska; the others require accreditation for a broader set of disciplines. National Science and Technology Council, *Strengthening the Forensic Sciences*, 5.

[448] The American Society of Crime Laboratory Directors voted to begin a voluntary accreditation program for their laboratories in 1981.

[449] Melson, "Crime Laboratory Accreditation: The Assurance of Quality in the Forensic Sciences," 1. Also see K. Melson, "Improving the Forensic Sciences through Crime Laboratory Accreditation," in *Wiley Encyclopedia of Forensic Science* (Wiley-Blackwell, 2009).

[450] Melson, "Crime Laboratory Accreditation: The Assurance of Quality in the Forensic Sciences," 1.

[451] National Research Council (NRC), *DNA Technology in Forensic Science*, 107.

[452] See U.S. Department of Justice, Office of the Inspector General,, *The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases*, Office of the Inspector General (Washington, DC, April 1997).

[453] American Bar Association (ABA), "Report of the ABA Criminal Justice Section's Ad Hoc Innocence Committee to Ensure the Integrity of the Criminal Process," in *Achieving Justice: Freeing the Innocent, Convicting the Guilty*, ed. P. C. Giannelli and M. Raeder (Chicago: American Bar Association, 2006).

Perhaps the most recognized recommendation for universal accreditation appeared in Recommendation 7 of the 2009 NRC report, which stated in unequivocal terms that "Laboratory accreditation and individual certification of forensic science professionals should be mandatory" and repeated later that "all laboratories and facilities (public or private) should be accredited" within a certain time period.[454] That recommendation led other national bodies to endorse universal laboratory accreditation. For example, the National Science and Technology Council, Committee on Science, Subcommittee on Forensic Science,[455] recognized that

> [i]mplementation of a quality management system, as required by ISO/IEC accreditation standards, is a sensible strategy to help decrease the likelihood of errors in testing results, data interpretation, and opinions. Properly implemented, forensic laboratory accreditation serves each of the core stakeholders in the criminal justice system—the prosecution, the defense, and the judiciary—and increases public trust in the criminal justice system.

Following the lead of the Subcommittee on Forensic Science, the NCFS issued a recommendation to the U.S. Attorney General to support universal accreditation of all Department of Justice forensic science laboratories, discussing both the benefits and challenges of accreditation. It concluded that "[u]niversal accreditation will improve [federal laboratory] ongoing compliance with industry best practices, promote standardization, and improve the quality of services provided by [federal laboratories] nationally."[456] The Attorney General adopted that recommendation.[457]

The accreditation process benefits forensic service providers in many ways.[458] Achieving laboratory accreditation is a means of assuring the technical competence of laboratories to perform specific types of testing, measurement, and calibration. It also gives formal recognition to laboratories that have taken the extra step of having their policies and procedures externally audited, providing customers with a level of confidence in the work being undertaken within those laboratories. The Working Group recognizes that accreditation guarantees neither the quality of a laboratory's work product and FDE competency, nor substitutes for validation. It does, however, provide several benefits:

- A series of benchmarks that define minimum requirements for quality documentation and generally accepted practices;
- An external and independent assessment of a service provider's management, technical, and quality policies, and checks if the policies are being followed;
- Formal recognition of meeting QA standards by an accreditation body;

[454] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*, 215.

[455] National Science and Technology Council, *Strengthening the Forensic Sciences*, 4.

[456] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: Universal Accreditation*, 2.

[457] U.S. Department of Justice, "Justice Department Announces New Accreditation Policies to Advance Forensic Science," news release, December 7, 2015, https://www.justice.gov/opa/pr/justice-department-announces-new-accreditation-policies-advance-forensic-science. Although the NCFS made recommendations to the Attorney General, it was seen as a leading policy body, speaking generally to the entire forensic science community. The same principles underlying its recommendation for federal laboratories apply to other laboratories as well.

[458] See *S. Bales, "Turning the Microscope Back on Forensic Scientists," Litigation 26, no. 2 (2000), https://www.jstor.org/stable/29760125.*

- Professional association with other experts from accredited providers (both nationally and internationally);
- External proficiency testing;
- A framework for a documented QMS; and
- Guidelines for ethical and professional responsibilities as outlined, for example, by the ANAB *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*.[459]

Depending on the region, accreditation for forensic service providers is offered by organizations like ANAB,[460] A2LA,[461] or the National Association of Testing Authorities (NATA).[462] Many of these accreditation organizations incorporate and build on the ISO/IEC International Standard 17025,[463] *General Requirements for the Competence of Testing and Calibration Laboratories*, by adding field-specific requirements.[464] The organizations use the same ISO/IEC 17025[465] standards regardless of the size of the laboratory. As noted above, some jurisdictions in the United States require accreditation of laboratories,[466] but historically, many forensic laboratories have become accredited voluntarily.

Although accreditation is a well-known and long-established component of a QMS in many laboratories, it poses challenges for small laboratories and private, sole practitioners in particular. The NCFS cited those challenges in its recommendation on universal accreditation.[467]

The NCFS, however, also presented suggestions to make the accreditation procedure less daunting for small and private laboratories. The NCFS noted that by implementing accreditation requirements in steps and in no required order, small laboratories could build toward an accreditation application rather than spending a significant amount of time and resources to do it all at once. The NCFS identified additional resources that may be of some assistance, like

---

[459] ANSI National Accreditation Board (ANAB). *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel*.

[460] "Forensic Accreditation," 2020, accessed May 6, 2020, https://anab.ansi.org/forensic-accreditation.

[461] "American Association for Laboratory Accreditation (A2LA) home page," 2020, accessed May 6, 2020, https://www.a2la.org.

[462] "National Association of Testing Authorities (NATA), Australia home page," 2019, accessed May 6, 2020, https://www.nata.com.au/.

[463] The ISO, a non-government international organization, creates voluntary, consensus-based international standards. ISO has partnered with its sister organization, IEC, which sets consensus-based international standards for electrical, electronic, and related technologies. Together, they have published standards for the competence of testing and calibration laboratories, known as ISO/IEC 17025. The current version of ISO/IEC 17025 was published in November 2017.

[464] Such as National Association of Testing Australia (NATA), *Specific Accreditation Criteria: ISO/IEC 17025 Application Document Legal (including Forensic Science)—Appendix* (2018), https://www.nata.com.au/phocadownload/spec-criteria-guidance/legal-forensic/Forensic-Science-ISO-IEC-17025-Appendix.pdf. and ANSI National Accreditation Board (ANAB). *ISO/IEC 17025:2005—Forensic Science Testing Laboratories Accreditation Requirements*.

[465] In 2017, an updated standard was published; however, the vast majority of crime laboratories in the United States are currently still accredited to the 2005 standard as there is a 3-year allotted transition period to fulfill any additional requirements of the 2017 standard. International Organization for Standardization (ISO), "New Edition of ISO/IEC 17025 Just Published," *ISO News*, December 1, 2017, https://www.iso.org/news/ref2250.html.

[466] As of 2013, 13 states and the District of Columbia had passed legislation mandating accreditation and other oversight requirements for at least some forensic service providers, including: Arkansas California, Hawaii, Indiana, Louisiana, Maryland, Missouri, Nebraska, New York, North Carolina, Oklahoma, and Texas. https://www.ncsl.org/Documents/cj/AccreditationOfForensicLaboratories.pdf; Accreditation is required only for laboratories conducting forensic DNA analysis in California, Hawaii, Indiana, and Nebraska; the others require accreditation for a broader set of disciplines. National Science and Technology Council, *Strengthening the Forensic Sciences*, 5.

[467] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: Universal Accreditation*, 2.

companies that provide training on quality management or the accreditation process and publicly shared documents on policies and procedures. It also recommended that small laboratories build networks through professional organizations or certification bodies to establish qualified reviewers and testing providers.[468]

Recognizing the many benefits of accreditation and the challenges inherent in achieving it, a majority of members of the Working Group were in favor of recommending that all forensic document examination laboratories should be accredited, whether they consist of a large team or a sole practitioner. This recommendation mirrors Recommendation 9.3.6 in the Latent Print report.[469]

A significant minority of members of the Working Group, that included all sole practitioners in private practice, did not support the accreditation recommendation. Although this group supports the goals of accreditation, they were troubled by several logistical shortcomings in its current implementation process. For example, it was noted that the checks and balances currently required for accreditation are designed to be undertaken by other designated persons. The minority expressed concern that civil litigation limits the FDE's ability to expose others to documents without violating confidentialities. Furthermore, it was noted that there were many instances in which the sole practitioner would wear multiple hats, essentially performing their own checks and balances. Although sole practitioners do perform checks and balances routinely, the types of checks and balances mandated by accrediting bodies are meaningful for a larger laboratory but not for a sole practitioner. This minority expressed a need to resolve the many implementation issues before recommending any accreditation requirements.

In addition to these practical constraints, the full Working Group recognizes that the accreditation process may be unnecessarily cumbersome, time-consuming, and costly regardless of laboratory size.

If the accreditation process could be carefully retooled to address the aforementioned concerns, the dissenting members of the Working Group stated they might be supportive of a recommendation for mandatory and universal accreditation. FDEs and associated professional organizations should collaborate with accrediting organizations to develop sector-specific requirements that address challenges for single FDE laboratories and private practitioners in addition to streamlining the overall process.

Recommendation 4.1a: Forensic document examiner laboratories* should be accredited to the current ISO/IEC 17025 standard by a recognized accrediting body.

***4.1b: In recognition of the practical constraints for sole practitioner laboratories to obtain accreditation, these laboratories should work toward meeting the requirements set forth in the current ISO/IEC 17025 standard and should become accredited when legitimate constraints are addressed.**

[468] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: Universal Accreditation*, 3.

[469] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*.

4.2. The QMS

This section explores the elements of QA and QC that sit within a QMS and their minimum requirements necessary for accreditation. A laboratory that has met accreditation requirements will already have these elements in place. However, it is the understanding of the Working Group that a significant number of FDEs do not work in externally accredited laboratories. A forensic service provider should develop a QMS regardless of whether the laboratory is accredited.

A healthy QMS will

- Strengthen competency—All FDEs must demonstrate competency before being allowed to examine casework and testify. Rigorous competency testing must include thorough analytical testing for all aspects of handwriting examinations and court training. See section 4.2.6.1.
- Maintain ongoing proficiency—The verification of ongoing FDE competency must be demonstrated. This is typically achieved by successfully undertaking at least one proficiency test every year. Testing through an accredited test provider is preferable. Proficiency tests must have known answers (i.e., ground truth), expected results, and provide feedback to the test taker. See section 4.2.6.2.
- Assist with laboratory accreditation—Laboratories should comply with international accreditation standards so that the overall "quality system" can be externally assessed for compliance with those standards.
- Regulate the review of policy and procedure manuals—Manuals should be reviewed at least biennially to ensure they are current and appropriate and so that policies and procedures can be refreshed in the minds of the FDEs and managers. See section 4.2.3.1.
- Regulate the review of examinations—Technical reviews of examinations are undertaken to help identify errors before issuing a report to the client. In addition, reviews can assist in monitoring and maintaining ongoing FDE proficiency. See section 4.2.3.2.

4.2.1. The Quality Manual

The backbone of a QMS is the quality manual, which is the source of the laboratory's policies and procedures. Many of the procedures described in the quality manual are applicable across disciplines in forensic science (for example, evidence handling), but issues specific to handwriting examination may be addressed where relevant.

The quality manual should document protocols to

- Define the organization, job duties, objectives, terminology, and abbreviations;
- Define staff educational and technical requirements;
- Establish and commit to a QMS;
- Establish and supervise the components of training and technical operations, focusing on quality laboratory results;

- Establish detailed, standardized methods for examinations and reporting;
- Establish requirements for documentation and review, which should document the frequency of review of the case records, reports, and testimony;
- Establish an approach toward errors that encourages transparency, appropriate root cause analysis, and corrective actions;
- Provide a guide to the proper management of the work environment and equipment;
- Provide procedures on how to handle records, evidence, and equipment appropriately;
- Ensure periodic audits of casework are conducted (both internal and external); and
- Enable continuous improvement of staff and their work output through training and certifications maintained through continuing education and other means.

If following ISO/IEC 17025, the above requirements in a QMS manual are divided into two primary sections: management and technical (covering resource and process requirements).[470] The management section of the quality manual addresses the role of management, whereas the technical sections focus on the resources and procedures surrounding the laboratory's work. The main areas that must be covered in these sections are summarized in table 4.1. Sections 4.2.2 through 4.2.8 highlight some of the technical requirements and activities of a QMS and how these may assist in reducing the negative impact of human factors on examinations. Discussion on human factors issues arising from the responsibilities of forensic handwriting laboratory management is covered more extensively in chapter 6.

Table 4.1: A summary of the key areas covered in the two main sections of a quality manual

Management	Technical
• The laboratory management's commitment to a code of professional ethics and to the quality of its testing and calibration in the services offered to its customers • The management's statement of the laboratory's standard of service • The purpose of the management system related to quality • The laboratory management's commitment to comply with the ISO standards and to continually improve the effectiveness of the management system	• Personnel (FDE qualifications, training, and competency, evaluations) • Accommodation and environment • Equipment • Test methods and their validation • Reports and reviews

(continued)

[470] Although ISO/IEC 17025:2005 has just the two primary sections, the requirement is upheld in the current ISO17025:2017 standard (International Organization for Standardization (ISO), "ISO/IEC 17025:2017(en) General Requirements for the Competence of Testing and Calibration Laboratories."); however, the format of the latter has been revised to follow the structure mandated by ISO/CASCO and as such is split into general, structural, resource, process, and management requirements. There is a 3-year allotted transition period to fulfill any additional requirements of the 2017 standard. International Organization for Standardization (ISO) New Edition of ISO/IEC 17025 Just Published.

Table 4.1: A summary of the key areas covered in the two main sections of a quality manual (continued)

Management	Technical
• The commitment of all personnel involved with testing and calibration activities within the laboratory to familiarize themselves with the quality manual and implement the policies and procedures in their work	

The quality manual establishes guidelines and expectations for all staff in the laboratory. This strengthens the QMS as a benchmark for maintaining work products, directing corrections when needed, and establishing a positive error culture that builds improvement into the current system.

Laboratory staff may write the quality manual, whereas non-technical content (like relating to management or general laboratory operations) may be established by the parent agency. The NCFS has recommended that all Department of Justice forensic science service providers, on request, make QMS documents accessible to the public in an electronic format.[471] Some laboratories already publish their quality manuals online, and these could be used as models for other laboratories developing their quality manuals or on the path to accreditation.[472]

Establishing and implementing a quality manual is a significant first step in the accreditation process. However, it cannot be considered as a replacement for accreditation because there are many additional benefits to accreditation like external assessment.

Recommendation 4.2: All forensic document examiner laboratories, whether accredited or not, must have a quality assurance and quality control system. This system should preferably align with the requirements of an international laboratory accreditation body.

4.2.2. Examination Methods/Procedures

Accredited laboratories are required to develop and maintain appropriate methods and procedures for the examinations performed. Documented methods and procedures benefit the laboratory system by providing guidance to FDEs for the steps expected in each examination. Although the QMS may suggest the format that best fits laboratory or accreditation expectations,

[471] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General Regarding Transparency of Quality Management System Documents*, Department of Justice (2016), https://www.justice.gov/ncfs/file/839706/download.

[472] See, for example: Indiana State Police Laboratory, *Quality Assurance Manual* (2016), https://www.in.gov/isp/labs/files/Lab_QA_Manual_03-16-16.pdf; Virginia Department of Forensic Science, *Quality Manual* (2017), https://www.dfs.virginia.gov/wp-content/uploads/2020/09/100-D100-DFS-Quality-Manual.pdf; Alaska Department of Public Safety Scientific Crime Detection Laboratory, "Alaska Crime Laboratory Quality Assurance Documents," (n.d.). https://dps.alaska.gov/Comm/CrimeLab/Quality-Assurance/QualityAssurance; "Quality Manuals," 2020, accessed May 6, 2020, https://www.crimelab.arkansas.gov/quality-manuals; "Welcome to Idaho State Police Forensic Services," updated April 13, 2020, https://isp.idaho.gov/forensics/; "Open Government and FOIA - DFS," n.d., accessed May 6, 2020, https://dfs.dc.gov/page/open-government-and-foia-dfs; Austin Police Department, Quality Assurance Section, *Standard Operating Procedures* (2016), https://web.archive.org/web/20170201010059/ https://www.austintexas.gov/sites/default/files/files/Police/QA_Standard_Operating_Procedures_01-11-16.pdf.

the procedures should follow field standards whenever possible. Laboratory policy should describe the steps to take if an examination deviates from the developed methods.

Implementation and use of standard operating procedures (SOPs) are critical to ensuring accurate and repeatable results for each type of analysis performed in handwriting examination. When laboratories developed operating procedures in the early days of forensic document examination, the procedures were typically based on a small number of highly regarded texts.[473] During the last quarter century, a more intense scrutiny of forensic document examination by the courts and critics has revealed that this forensic discipline has lacked specific and universally accepted research-based standards for the work performed by FDEs. These criticisms spurred the development of a series of standards and formalized processes.

The NIJ and FBI began developing standards for the field of forensic document examination in 1997. The website for the SWGDOC[474] describes the organization and its history. SWGDOC is composed of private and government FDEs from local, state, and federal laboratories throughout the United States, with additional international representation of FDEs. SWGDOC began in 1997 as the Technical Working Group for Questioned Documents, was renamed SWGDOC in 1999, and was reorganized in 2001. From 2000 to 2012, SWGDOC-drafted standards were reviewed, revised, and published through ASTM.

In 2012, SWGDOC began self-publishing its standards like other scientific working groups. In 2014, OSAC took on the task of creating and reviewing standards in preparation for the standards development organization process. The American Academy of Forensic Sciences (AAFS) established the Academy Standards Board (ASB) in 2015 and obtained accreditation from ANSI. OSAC's forensic document examination subcommittee will submit its revised standards (based largely on what has been produced by SWGDOC) to ASB for vetting and for establishing national standards in the field.

The Working Group suggests that standards are based on empirical data to support the claims made by FDEs regarding the reliability and validity of forensic handwriting examination. (See section 2.3, outlining important research needs and section 2.2, dealing with validity and reliability of forensic handwriting examinations.) Once consensus standards (such as those being produced by OSAC) are developed and approved, their adoption has the potential to assist FDEs in recalling and following all steps in the examination process, streamlining the review procedure, and explaining the examination process to external reviewers and customers.

Given the concerns about contextual bias in forensic examinations (see section 2.1) the QMS should assist in setting laboratory policies to facilitate appropriate CIM procedures for handwriting examination, whenever possible. This documentation should include definitions of task-relevant versus task-irrelevant information. [475]

---

[473] Such as A. S. Osborn, *The Problem of Proof: Especially as Exemplified in Disputed Document Trials: A Discussion of the Proof of the Facts in Courts of Law: With Some General Comments on the Conduct of Trials* (Newark, NJ: Essex Press, 1926); Osborn, *Questioned Documents*; Harrison, *Suspect Documents: Their Scientific Examination*. These texts were followed by James V. P. Conway, "Evidential Documents," in *Scientific Examination of Questioned Documents*, ed. Ordway Hilton (CRC Press, 1959).

[474] "Scientific Working Group for Forensic Document Examination (SWGDOC) home page," updated March 20, 2018, 2019, https://www.swgdoc.org.

[475] National Commission on Forensic Science (NCFS), *Views of the Commission: Ensuring that Forensic Analysis Is Based Upon Task-Relevant Information*.

4.2.3. Review

An accredited QMS offers many levels of review. Each level improves feedback to personnel and management in a distinctive way. Reviews may be external reviews through accreditation, internal audits that include review of laboratory management policies and procedures, and casework reviews with corresponding policies regarding nonconforming work.

4.2.3.1. Internal Audits

An internal audit of management system documents and a review of these documents are separate processes, but both work toward similar goals. Conducted between reassessment visits by external auditors, both processes are directed internally and focus on staff, safety, and maintenance with respect to requirements specified under the accreditation rules. Records of the findings of any audit, and changes implemented because of the process, must be maintained, and contribute to the quality system's overall documentation. Some accreditation programs also require an annual review of ethics guidelines, which can also be accomplished during these internal reviews.

Other considerations for sole practitioner or small laboratory

Technical reviews for a sole practitioner, whether in private practice or part of a larger laboratory, may present challenges, including

- In particularly sensitive cases, the submitter may not want the documents to be reviewed by another person;
- Suitable reviewers may be difficult to locate and engage with in a timely manner; and
- The potential associated cost consideration may add to the cost of the examination for the client.

However, from a human factors perspective, the benefits of technical review may outweigh these challenges.

4.2.3.2. Casework Reviews

Casework reviews serve as a critical part of the QMS. Casework reviews serve as a key mechanism for ensuring the "accuracy and completeness of the opinion and associated documentation."[476] The range of casework review types—administrative, technical, and re-examination—differ in the level of scrutiny they offer and the technical background of the reviewer. Casework review builds a level of redundancy into the system and serves as a tool for improving overall system quality. Redundancy within the system does not render the conclusion infallible, but it can serve as a reliable way to detect and ultimately reduce the number of errors that leave the system. Although agency policies vary in how casework reviews are undertaken, two common elements are (1) the reviews should be conducted by someone other than the assigned FDE and (2) in the interest of transparency, the identity of the reviewer(s) should be documented.

Administrative Review

An administrative review examines the case file and report to ensure that the relevant case work or quality systems procedures have been followed (evidenced via inclusion of appropriate documentation in the case file)[477] and to check the use of correct grammar and spelling. An administrative review also checks that the final report is coherent and reflects the examinations

[476] K. N. Ballantyne, G. Edmond, and B. Found, "Peer Review in Forensic Science," *Forensic Science International* 277 (Aug 2017), https://doi.org/10.1016/j.forsciint.2017.05.020, https://www.ncbi.nlm.nih.gov/pubmed/28622536.

[477] Queensland Police Service, *Forensic Services Quality Manual* (2015).

performed and the reporting requirements.[478] It is acceptable for administrative reviews to be undertaken by someone outside of the area of expertise but who is familiar with the laboratory's QMS.

Technical Review

A technical reviewer examines the case file (bench notes, data, and other documents that form the basis for scientific conclusions[479]) to ensure the reported conclusions fall within the scope of the discipline and applicable policies and are supported by sufficient data.[480] This kind of review does not usually (although it can) involve the full re-examination of the evidence but is a precaution taken to ensure that the correct and appropriate procedures have been followed and documented, that the conclusions reached are supported by the observations, and that the results are documented in the case file.[481] Technical reviews must therefore be carried out by someone who is qualified in the relevant discipline. It is acceptable for administrative and technical reviews to be performed as part of one review process.[482] The Working Group suggests that organizations have a checklist or worksheet so that a reviewer can identify and understand the elements of the review process. Although technical reviews are an important aspect of a laboratory's QMS, they should not be used to shift the perceived responsibility for the scientific findings from the FDE to the reviewer. It is the FDE who issues the report and presents testimony regarding the findings.[483]

ISO/IEC 17025 currently requires that laboratory results be reviewed and authorized before release.[484] One forensic science accreditation body[485] makes it explicit that 100% of case files must be technically and administratively reviewed unless the risk associated with undertaking fewer reviews has been calculated. Some laboratories choose to only conduct technical reviews on certain case types or for certain results. For example, a laboratory may only conduct technical reviews on cases where an association was made.

The Working Group believes a mixture of cases, including where testimony is anticipated, should undergo technical review. Also including cases where testimony is not required will help ensure that the process is sufficiently blinded. Reviewing these cases would increase the chance

---

[478] See New York City Office of Chief Medical Examiner, *Forensic Biology Evidence and Case Management Manual* (2019), https://www1.nyc.gov/assets/ocme/downloads/pdf/technical-manuals/forensic-biology-evidence-and-case-management-manual/administrative-review.pdf; North Carolina State Crime Laboratory, *Procedure for Reviewing Laboratory Reports* (2013), https://forensicresources.org/wp-content/uploads/2019/07/Reviewing-Laboratory-Reports-08-29-2014.pdf; National Association of Testing Australia (NATA), *General Accreditation Criteria: Legal Management of Facility Activities (Forensic Operations Module)* (2019), 7, https://www.nata.com.au/phocadownload/gen-accreditation-criteria/Forensic-Operations-Module.pdf; J. M. Taupin, *Introduction to Forensic DNA Evidence for Criminal Justice Professionals* (Boca Raton: CRC Press, 2013), 61.

[479] U.S. Department of Justice, Office of Justice Programs, *Forensic Sciences: Review of Status and Needs* (February 1999).

[480] National Association of Testing Australia (NATA), *Specific Accreditation Criteria: ISO/IEC 17025 Application Document Legal (including Forensic Science)—Appendix*, 17.

[481] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*.

[482] National Association of Testing Australia (NATA), *Specific Accreditation Criteria: ISO/IEC 17025 Application Document Legal (including Forensic Science)—Appendix*, 17.

[483] National Association of Testing Australia (NATA), *Specific Accreditation Criteria: ISO/IEC 17025 Application Document Legal (including Forensic Science)—Appendix*.

[484] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*. Section 7.8.1.1.

[485] National Association of Testing Australia (NATA), *Specific Accreditation Criteria: ISO/IEC 17025 Application Document Legal (including Forensic Science)—Appendix*, 17.

for detecting and correcting a technical error before testimony. It is the understanding of the Working Group that many accredited government forensic handwriting laboratories conduct a technical review of all their cases and that the reviewer must agree with the opinions of the FDE (within a certain tolerance) before a report is issued. In these instances, the technical reviewer may do more than merely check that the opinion is supported by the documentation.

Although some research has demonstrated that the reliability of forensic document examination is increased by technical review,[486] there is also some concern that the nature of current technical review processes is not adequate to achieve the desired aims of the review (i.e., to reduce the potential for errors in the application of procedures and in opinions).[487] For the error potential to be reduced, some level of reanalysis is required.

Re-examination

Re-examination occurs when two or more FDEs independently examine and evaluate the same material and form their own conclusions. Re-examination of casework can be non-blinded or blinded. In a non-blinded re-examination, a second FDE performs a full examination of either all the items submitted in the case or may be restricted to only examining the evidence items on which the initial FDE relied.[488] The reviewer is aware that an initial examination was conducted and is asked to reach and document their own conclusions. The reviewer may have access to the case notes, reports, and identity of the initial FDE. This type of review may also be referred to as verification.

In a blinded re-examination, a second FDE performs a full independent examination not knowing what the first FDE did or concluded and focuses completely on the evidence and comparisons.[489] The second FDE may or may not be blinded to task-irrelevant contextual information. If the second FDE is unaware that an initial examination was performed, this becomes a double-blind re-examination. The second FDE's findings/conclusions are documented. This approach—sometimes referred to as blind verification—combats the base rate expectation that arises from reviewing only certain opinion results.

Laboratory casework review policies vary widely as do the terms used to describe the three casework review types: administrative and technical reviews and re-examination. Within a forensic laboratory setting, one or more of these casework review types may sometimes be referred to as peer review.[490] However, the term peer review is more widely used to describe the process of reviewing manuscripts submitted for publication to a scientific journal. Analogously, in this context, one or more members of the relevant scientific community critically evaluate the

[486] T. Y. Kang and J. Lee, "Multiform Korean handwriting authentication by forensic document examiners," *Forensic Science International* 255 (Oct 2015), https://doi.org/10.1016/j.forsciint.2015.07.002; M. Durina and M. Caligiuri, "The Determination of Authorship from a Homogenous Group of Writers," *Journal of the American Society of Questioned Document Examiners* 12, no. 2 (2009).

[487] Ballantyne, Edmond, and Found, "Peer Review in Forensic Science."

[488] A lesser form of re-examination is based on copies of the items that the initial FDE replied on, rather than the same material that the initial FDE viewed.

[489] Itiel E. Dror, "Practical Solutions to Cognitive and Human Factor Challenges in Forensic Science," *Forensic Science Policy & Management: An International Journal* 4, no. 3-4 (2014), https://doi.org/10.1080/19409044.2014.901437; N. K. P. Osborne and M. C. Taylor, "Contextual Information Management: An Example of Independent-Checking in the Review of Laboratory-Based Bloodstain Pattern Analysis," *Science & Justice: Journal of the Forensic Science Society* 58, no. 3 (May 2018), https://doi.org/10.1016/j.scijus.2018.01.001.

[490] See, for example, Triplett and Cooney, "Etiology of ACE-V and Its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing."; Ballantyne, Edmond, and Found, "Peer Review in Forensic Science."

results presented, which acts as a form of QC.[491] Peer review of submitted scientific manuscripts can be single-blind in that the reviewer does not know the identity of the authors or can sometimes be double-blind, in which case neither the reviewer nor the authors know each other's identity. The Working Group avoids the use of peer review to refer to case work review in this report but recognizes that it has been used frequently in forensic science.

4.2.3.3. Nonconforming Work

If a case record review reveals nonconforming work, the QMS must address it quickly and appropriately. Nonconforming work may include problems associated with deviation from procedures, or improper interpretation or conclusions. The quality manual should include clear policy and definitions for the resolution of technical variations, conflicts of opinion, and nonconforming work. The process may use a panel of FDEs, a technical leader, or rely on outside consultation. The goal should be to set a standard for when and how the discovery of nonconforming work is reported to the customer. Documentation and transparency of the conflict and its resolution should be extensive, regardless of the results. A corrective action review, covered further in section 4.2.5, is intended to identify the cause of the nonconforming work, how to address and resolve the situation, and how to prevent the situation reoccurring in the future.

4.2.3.4. Human Factors Issues with Reviews

Although these review processes are designed to detect variations in product quality, noncompliance with procedures, or error, they may also be subject to human error. Particular care must be taken to minimize potential bias arising from the technical or administrative review process. For example, a preferred coworker may be consulted for review or a pair of reviewers may build a relationship to minimize turnaround time. Although these types of arrangements may have developed with good intentions, they can result in unconscious base rate expectation bias—an expectation that the technical and/or administrative components of the case will be adequate or that because of perceived competence, the result will be correct. To mitigate such biases, reviewers should be regularly changed and randomly selected from a pool of qualified FDEs whenever possible.

Compounding this expectancy problem is the pressure for reviewer and FDE to agree—perhaps due to their relationship or the culture of the laboratory, particularly in regard to conflict resolution and error management.[492] Selection of a casework reviewer must therefore consider any hierarchical structure that may exist. The most obvious human factor issue associated with hierarchy occurs when an individual perceived to hold greater power (either because of their position in a management hierarchy or by virtue of experience) provides a case to a lower ranking or less experienced individual for technical or administrative review.[493] The potential for

[491] See, for example, A. W. Jones, "The Distribution of Forensic Journals, Reflections on Authorship Practices, Peer-Review and Role of the Impact Factor," *Forensic Science International* 165, no. 2-3 (Jan 17 2007), https://doi.org/10.1016/j.forsciint.2006.05.013; J. L. Mnookin et al., "The Need for a Research Culture in the Forensic Sciences," *UCLA Law Review* 58, no. 3 (2011), https://heinonline.org/HOL/LandingPage?handle=hein.journals/uclalr58&div=20&id=&page=.

[492] Dror, "Cognitive Neuroscience in Forensic Science: Understanding and Utilizing the Human Element."

[493] See, for example, "trans-cockpit authority gradient," where flight crew pairing of very senior flight captains with junior co-pilots is likely to result in problems in communication and coordination. S. A. Shappell and D. A. Wiegmann, *The Human Factors Analysis and Classification System–HFACS*, Office of Aviation Medicine (Washington, DC, 2000), 10, https://www.nifc.gov/fireInfo/fireInfo_documents/humanfactors_classAnly.pdf.

bias is difficult to control under these circumstances, but one solution could be to ensure that reviewers are blinded to the conclusions, allowing them to reach an opinion based on the evidence before reviewing the full case file and report.

The knowledge that a case file will be reviewed may also be associated with differences in human reaction. Some FDEs, knowing their work will be checked by someone else, may take less care in their work. Other FDEs with the same knowledge may take extra care.

4.2.4. Monitoring of Results and Testimony

FDEs usually complete their examination by writing a report of the results and sometimes providing accompanying testimony for the judicial system. The QMS should monitor these products as they directly reflect on the FDE, the laboratory, and the practice.

The QMS should ensure that the report is accurate, unambiguous, and impartial; meets accreditation/laboratory policies; and that the release of the report to the customer is documented (see chapter 3 for further information regarding report writing).[494] It may be helpful for the laboratory to understand how the client uses and interprets the report.

Because expert testimony could be a critical part of a court case, the QMS should have policies in place to review the performance of those testifying. Accreditation by ANAB mandates that each FDE receive training in professional ethics and "criminal law, civil law, and testimony"[495] and that the FDE's testimony be monitored at least once per year. This monitoring may be carried out by direct observations (recorded on an evaluation form), review of transcripts, or telephone solicitation.

The evaluation should consider the FDE's behavior on the stand, including appearance, poise, and performance under direct and cross-examination. For example, if the FDE pauses longer between the question and answer on cross-examination than on direct examination or adopts a much more rigid facial expression or posture, the fact finder may construe that as evidence of an underlying bias, and it could undermine the credibility of the FDE's testimony. This same concept applies to testimony at a videotaped deposition.

Similar problems may arise if the FDE is repeatedly nonresponsive on cross-examination, which may allow an opposing attorney to undermine testimony on the basis of perceived poor or hostile conduct. In addition to behavior, the evaluation should also assess the FDE's communication skills. The evaluation should determine whether the FDE has the ability to present evidence so that the judge and jury can understand the material and whether the FDE's testimony is consistent with the case records and report and does not overstate the findings. Relevant research should include how the FDE's presentation of evidence in court impacts the judge and jury's comprehension of the forensic evidence to avoid potential misunderstandings or miscommunication.[496]

---

[494] See also National Commission on Forensic Science (NCFS), *Views of the Commission: Documentation, Case Record and Report Contents*.

[495] ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories*, 7.

[496] K. Browning, "Social Science Research on Forensic Science: The Story Behind One of NIJ's Newest Research Portfolios," *NIJ Journal* 275 (2015), https://nij.gov/journals/275/Pages/social-science-forensics.aspx#.

The QMS establishes policies specifying actions that should be taken for negative or critical evaluations. Monitoring also gives the manager additional information with which to evaluate employees where relevant and may reveal that some FDEs need more practice (e.g., by participating in mock cases), training, and feedback than is currently given to develop adequate testimony and presentation skills. Feedback from the court system regarding testimony could also be useful for improvements to the overall laboratory system.

Data show that more than 90% of criminal cases are settled through plea negotiations.[497] If the report is the only document available to those negotiating the plea, then it carries significant weight on the outcome but does not face the same scrutiny of courtroom proceedings as testimony. These concerns could also extend to other stages of processing, such as changing decisions and alternative dispute resolution that may occur outside of court records available to the public.[498]

Further discussion on human factors issues relating to testimony and recommendations to mitigate these issues can be found in sections 3.4 to 3.6.

4.2.5. Preventive and Corrective Actions

Corrective actions and preventive actions are additional components of a QMS. In terms of QA, policies and procedures will provide for implementation of preventive actions whereas corrective actions are QC for nonconforming work, whether in relation to technical or management requirements.[499]

When nonconforming work is detected or reported, a corrective action policy first assesses the nonconformity's significance regarding the potential impact and actual risk to the evidence, analysis, or work product. Some laboratories classify the nonconformity into a level, class, or type of error with definitions and approaches to determine the course of action. For example, a laboratory's QMS may define a Level 1 nonconformity as unexpected and an immediate concern regarding the quality of the work or integrity of the evidence.[500] Furthermore, Level 1 requires investigation into a root cause by more than one individual and extensive corrective actions with ample documentation. A root cause analysis should focus on implementing change to avoid future recurrence, enabling the laboratory to learn from the nonconformity, and allowing for a blame-free analysis with discipline issues handled in a separate process.[501]

If a full corrective action review takes place (i.e., for a "Level 1" nonconformance), the root cause, recommended course of action, and schedule to correct or follow-up should be outlined and distributed to the appropriate parties. An announcement to parties like the laboratory,

[497] J. Butler, *Advanced Topics in Forensic DNA Typing: Interpretation* (Oxford, UK: Academic Press, 2015), 3.

[498] D. McClure, *Focus Group on Scientific and Forensic Evidence in the Courtroom*, National Institute of Justice (Washington, DC, 2007), 11, https://www.ncjrs.gov/pdffiles1/nij/grants/220692.pdf.

[499] Indiana State Police Laboratory, *Quality Assurance Manual*, 29.

[500] For example, see *Non Conformity, Corrective and Preventive Action Procedure,* Procedure No. QP08.0007, (Harris County Institute of Forensic Sciences, 2016), 4.

[501] National Commission on Forensic Science (NCFS), *Directive Recommendation: Root Cause Analysis (RCA) in Forensic Science*, Department of Justice (2015), https://www.justice.gov/archives/ncfs/file/786581/download.

accreditation body, customers, and others associated with the case outside of the laboratory may be required. This could be covered by a duty-to-correct or duty-to-notify policy (see box 4.1).

Box 4.1: Duty to correct

FDEs duties do not begin and end with their report or testimony. Rather, an FDE must provide appropriate information before, during, and after trial. Indeed, there is "an ethical obligation to 'take appropriate action if there is potential for, or there has been, a miscarriage of justice due to circumstances that have come to light, incompetent practice or malpractice'."[502] Just as it is not the FDE's role to determine guilt or innocence (or liability or lack of liability, in civil matters), it is also not their role to determine whether a "miscarriage of justice" has occurred. Instead, the FDE has a responsibility before and during trial to ensure that the information provided is scientifically appropriate and conveyed in a competent and accurate manner. However, there may be instances where a report is retrospectively found to be based on unsound science or to involve incompetent practice or malpractice. In those instances, the FDE should report the matter to management for additional review. If the laboratory determines that previously offered testimony has the potential for or has caused a miscarriage of justice, the laboratory has a responsibility to take appropriate action. For FDEs in sole or small group practices who practice without laboratory managers, the FDE should notify the relevant attorneys.

Appropriate action may depend upon the jurisdiction in which the expert testified or for which the report was prepared and the policy of the FDE's laboratory. For example, in September 2016, the Attorney General approved a Code of Professional Responsibility for the Practice of Forensic Science for Department of Justice laboratories. Paragraph 16 states that the forensic science service provider management must "[i]nform the prosecutors involved through proper laboratory management channels of material nonconformities or breaches of law or professional standards that adversely affect a previously issued report or testimony." Nonconformities are defined in the Code as any "aspect of laboratory work that does not conform to its established procedures. An evaluation of the nonconformity risk is appropriate to deciding whether or not reporting is necessary."[503]

The NCFS recommends "all forensic science and forensic medicine service providers, associated certification and accreditation bodies, and professional societies to adopt the [Code], and for their management systems to develop policies and procedures to enforce the standards embodied in this code."[504] Testimony may be fully in line with a laboratory's protocols, the relevant laws, and professional standards at the time it is given, but the appropriateness and value of testimony shift as science evolves and as those parameters change in response. Put differently, bad faith, incompetence, and malfeasance are not required to trigger the need for a correction, so the duty to correct must be understood broadly.

[502] American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB), *Potential Issues with Hair Comparison Testimony: Notification from the ASCLD/LAB Board of Directors to Interested Parties Concerning Potential Issues with Hair Comparison Testimony* (2013).

[503] Attorney General, *Memorandum for Heads of Department Components: Recommendation of the National Commission on Forensic Science; Announcement for NSFS Meeting Eleven*.

[504] National Commission on Forensic Science (NCFS), *Recommendation to the Attorney General: National Code of Professional Responsibility for Forensic Science and Forensic Medicine Service Providers*, 4.

Moreover, given that many FDEs practice outside of large laboratories, the Working Group believes that professional societies have an important role in encouraging and supporting the duty to correct by FDEs. Accordingly, professional societies should consider adopting a duty to correct as part of their codes of conduct.

The Working Group acknowledges that implementing the duty to correct may differ between laboratories, because in some laboratories, issues can be reported upwards internally before a decision to report outwards (or not) is made. Although challenges may exist for the sole practitioner who has no management chain, there is still an obligation to correct materially inappropriate testimony, particularly in criminal cases. Such a process may involve notifying the relevant attorney of that issue and, if the FDE believes that the error affected other cases, a review of that testimony as well.

A Level 2 nonconformity in the same laboratory[505] is a minor deviation from policy or procedures, addressed as part of routine business, that may compromise the quality of the work product but is not persistent or serious enough to cause immediate concern. Level 2 nonconformities can be addressed by a single individual, consultation, or retraining with appropriate documentation.[506]

If the potential for nonconformity is reported, then a preventive action is put into place instead. Just like other reviews, preventive actions should be addressed appropriately, reviewed with staff, and documented.

Human factors play a key role in corrective and preventive actions within the QMS. The QMS should not only anticipate potential error but should also have procedures in place for how to deal with error(s) and then improve the system to minimize recurrence. More importantly, forensic science requires a culture in which the impact of nonconforming work is addressed openly and promptly. A clear policy should be communicated to employees about the results of corrective actions so that termination is not feared when retraining would suffice. See section 6.3 for a discussion on positive error culture.

Some avenues for reporting nonconforming work include reports by employees or customers about other employees or about themselves. If the employee is afraid, discouraged, or otherwise prevented from reporting nonconforming work, the entire system suffers. Additionally, corrective and preventive actions need oversight by employees with the management authority.

Because of the nature of forensic work and the fact that life and liberty may depend on the accuracy of laboratory results, corrective and preventive actions should be part of any QMS. There may be instances when independent, external FDEs are called in to investigate cases of suspected negligence, misconduct, or systemic misapplication of forensic science.[507]

---

[505] Harris County Institute of Forensic Sciences. *Non Conformity, Corrective and Preventive Action Procedure*, 5.

[506] Virginia Department of Forensic Science, *Quality Manual*, 32, 34.

[507] "Overturning Wrongful Convictions Involving Misapplied Forensics," n.d., accessed May 6, 2020, https://www.innocenceproject.org/causes/misapplication-forensic-science/.

4.2.6. Personnel and Laboratory Testing

Within a QMS, two types of ground truth tests are encountered: competency tests and proficiency tests. These are described and discussed in sections 4.2.6.1 and 4.2.6.2. Ground truth tests that are not generally discussed within a QMS include collaborative, black box, white box, and blind declared cases, but they are referred to in the context of establishing validity. See section 2.2.2 for discussion regarding black box and white box studies.

4.2.6.1. Competency Testing

The purpose of a competency test is to determine whether a forensic science practitioner has acquired and can demonstrate specialized technical knowledge, skills, and abilities in the standard practices relating to examinations in a specific discipline or category of testing. Competency testing is an integral part of the forensic training process and must be successfully completed before performing independent casework. Competency testing may take the form of written, oral, practical, or role exercise (e.g., mock court) tests.[508] This kind of testing does not assess a forensic service provider's overall quality system and performance (including methods, procedures, testimony, reports, documentation, equipment, validation, measurement uncertainty, facilities, evidence handling, security, or safety procedures used by the individual practitioner[509]) but does evaluate an FDE's ability to reach appropriate conclusions in the tested area. Further considerations regarding an FDEs' competence are discussed in section 5.3 and section 6.2.

4.2.6.2. Proficiency Testing

In an accreditation environment, the term proficiency test has a specific meaning. It is a recognized QC tool designed to evaluate participant performance against pre-established criteria by means of inter-laboratory comparisons.[510] Proficiency testing evaluates the performance of individual laboratories based on specific tests or measurements. The testing also monitors the continuing performance[511] and quality system of laboratories and their ability to adhere to the organization's documented procedures.[512]

The first step in the process is the actual testing and identification of any errors made with a follow-up step to try to identify the root cause of errors and initiate action for improvement/correction. In this way, proficiency testing allows a laboratory to discover systemic issues[513] (for example, in procedures, environment, training, or calibration of equipment) by monitoring a laboratory's long-term performance, and those issues can then be investigated and corrected.

Although proficiency tests alone are not suitable for assessing an FDE's competence upon completion of training, these tests are used to monitor an individual FDE's continued ability to

---

[508] European Network of Forensic Science Institutes (ENFSI), *Best Practice Manual for the Forensic Examination of Handwriting*, 5.

[509] National Commission on Forensic Science (NCFS), *Proficiency Testing in Forensic Science*, Department of Justice (2016), https://www.justice.gov/archives/ncfs/page/file/831806/download.

[510] *Conformity Assessment—General Requirements for Proficiency Testing,* ISO/IEC 17043:2010, (Geneva, Switzerland: International Organization for Standardization (ISO), 2010).

[511] "What is Proficiency Testing?," updated October 24, 2019, 2001, https://www.hn-proficiency.com/profi.htm.

[512] National Commission on Forensic Science (NCFS), *Proficiency Testing in Forensic Science*.

[513] College of Physicians and Surgeons of British Columbia, *Laboratory Medicine Proficiency Testing Manual*, Diagnostic Accreditation Program (2018), 3, https://www.cpsbc.ca/files/pdf/DAP-PT-Manual.pdf.

perform specific tasks or work within a specific discipline. The use of proficiency testing in this manner should not be confused with competency testing.[514]

Proficiency tests may also be able to

- Establish the effectiveness and comparability of test or measurement methods;
- Identify inter-laboratory differences;
- Provide feedback to participating laboratories based on the outcomes of such comparisons; and
- Validate uncertainty claims.[515]

These tests are generated by registered proficiency test providers for use as part of the accreditation process for laboratories. However, the NCFS took a broader view of proficiency tests as a valuable tool, regardless of whether they are used for accreditation.[516] At present, there is only one accredited, English-based proficiency test provider for handwriting examinations.[517] Typically, the tests emulate the circumstances and materials that might be expected of routine casework. These proficiency tests may be focused on handwritten uppercase, lowercase, or printed material, signatures, or a combination of these.

Limitations Associated with Proficiency Testing

Proficiency tests are valuable because the ground truth is known, and practitioners are provided with feedback about whether their results concur with the manufacturer's results. Because results are provided to participating laboratories and practitioners, practitioners also have the opportunity to compare performance with other test takers. However, proficiency tests have two major limitations.

First, a proficiency test does not provide information on when an inconclusive opinion regarding writership is the most appropriate opinion for an FDE to give. For instance, although casework often comprises far more complex writing, there are occasionally comparisons that involve fewer characters like truncated signatures, initials, or other abbreviated text. To illustrate this point further, consider an extreme example of a questioned single sans serif numeral 1 (i.e., a single vertical line). The ground truth of this single numeral 1 is that it was written by the writer of the known handwriting sample. If the known handwriting sample contains a substantial number of sans serif numeral 1s, an FDE expressing the opinion that the questioned 1 was written by the known writer would be correct with respect to the ground truth. However, it would be negligent to not also express that it could be equally likely that someone other than the comparison writer wrote the single stroke (and therefore that no opinion can be expressed regarding writership). Furthermore, even though opinions can be compared with the consensus results of other participants, it may be that they are not an appropriate group to compare against because the

[514] National Commission on Forensic Science (NCFS), *Proficiency Testing in Forensic Science*.

[515] International Organization for Standardization (ISO). *Conformity Assessment—General Requirements for Proficiency Testing*.

[516] National Commission on Forensic Science (NCFS), *Proficiency Testing in Forensic Science*..

[517] "Collaborative Testing Services (CTS), Inc. home page," 2020, accessed May 6, 2020, https://cts-forensics.com/.

nature of these other test takers is unknown (e.g., trainees or experts trained but following a different test procedure).

Second, because proficiency tests are based on typical casework, the test provides only limited information even if completed successfully. Participants will know whether their results agree with the manufacturer's known answer, but they will not know whether their results are correct for the right reasons. Suppose that an FDE determines the questioned signatures are genuine. If only genuine signatures were presented in the test material and an FDE were to opine that the questioned signatures were genuine, the individual would be correct and pass the proficiency test without knowing their proficiency in assessing simulated signatures.

Now imagine providing the same test but with one of the questioned signatures simulated. In this case, if the FDE opined that all of the questioned signatures were genuine, they would be correct for questioned genuine signatures but in error for the questioned simulated signature. This provides meaningful feedback for the claim that the practitioner is proficient in discriminating between genuine and simulated signatures. The composition of the questioned population (e.g., genuine, disguised, and simulated) affects the value of the test.

The challenge is to develop tests that demonstrate something about the FDE's proficiency while also reflecting casework. In typical casework samples, it is unlikely there will be alternate proposition questioned samples representing the range of claims that FDEs make (e.g., genuine, disguised, simulated). Proficiency tests are therefore limited to the extent to which they inform an FDE's proficiency unless they show error. In addition, the test materials alone cannot be used to demonstrate task validation.

Proficiency test design can also impact an FDE's responses to the test. A 2017 analysis of 10 years of proficiency test data from Australian government forensic service providers highlighted that from 2005 to 2015, one handwriting proficiency test was designed differently than previous tests, which all followed a familiar pattern. This change in design affected 4.71% of results (reportedly because of expectation bias).[518] A review by a Working Group member of CTS Summary Reports from 2007–2017 found that all of the questioned handwriting was naturally written, whether by one of the known writers or an individual whose known handwriting was not provided to participants. Questioned signatures fell into one of three categories: naturally written (by a known writer or someone else), disguised (specifically, the writer instructed to produce a simplistic wavy or looped line signature to not provide enough characteristics for identification), or signatures produced by known writers instructed to sign in a different name.[519] In only 2 of the 10 tests was there more than one contributor to the questioned handwriting (excluding signatures) on a single document. None of the tests contained disguised or simulated handwriting or simulated signatures. Cases with more than one contributor to the questioned writing or containing unnatural writings can be expected to be more complex and potentially ambiguous, but these scenarios are typically not encountered in proficiency tests.

Consideration should be given to assessing the frequency of testing, because even if the tests are given often enough to meet accreditation requirements, the frequency may not suffice to provide

[518] Linzi Wilson-Wilde, Stephen Smith, and Eva Bruenisholz, "The Analysis of Australian Proficiency Test Data over a Ten-Year Period," *Forensic Science Policy & Management: An International Journal* 8, no. 1-2 (2017), https://doi.org/10.1080/19409044.2017.1352054.

[519] Note that no model signatures were provided so this cannot be considered a simulation.

meaningful feedback on the full array of expertise claims that practitioners make regarding casework for their particular laboratory. In addition, even though proficiency tests are supposed to be carried out according to laboratory protocols, the tests are generally known to FDEs (i.e., it is obvious that the case examined is a proficiency test); therefore, the conclusions they reach may not accurately reflect performance in normal practice.[520] For example, extra attention may be given to the process or additional tests applied to the case samples to be sure of reaching a correct conclusion. Injecting these tests into the normal case flow would be challenging under normal laboratory processes (see section 4.2.6.6).

If an FDE's responses do not fit the manufacturer's report or are not in consensus with other responses, significant action may be undertaken. This action may include a corrective action review, reporting to the accreditation body, and other follow-through actions based on the root cause analysis.

Proficiency tests are generally not useful for testing the limits of FDE expertise when they are faced with difficult cases or ambiguous evidence,[521] which may be the cases most vulnerable to error. Although proficiency tests provide a ground-truth-known experience for practitioners and play an important role in the QMS, the Working Group is concerned practitioners may view proficiency tests as a means to support all FDE claims of expertise.

Additional recommendations and guidelines for proficiency testing can be found in the NCFS's *Views of the Commission Regarding Proficiency Testing in Forensic Science*.[522]

Recommendation 4.3: The forensic document examiner community should collaborate with the research community and accreditation bodies to conduct and participate in studies to determine the optimal content and frequency of proficiency tests to properly evaluate forensic document examiners' ability to perform the range of tasks encountered in casework.

4.2.6.3. Collaborative Testing

In a forensic context, collaborative testing refers to inter-laboratory trials, in which several laboratories examine the same material (either exactly the same material passed from one laboratory to the next [round robin] or duplicate material sent to each laboratory).

---

[520] NCFS (National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*, Department of Justice (2016), https://www.justice.gov/archives/ncfs/page/file/909311/download.) notes the following: "Informing someone that he or she is being tested can create what psychologists call demand characteristics that change the person's responses. Martin T. Orne, "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist* 17, no. 11 (1962), https://doi.org/10.1037/h0043424. Individuals who know they are being tested may shift their threshold of decision in ways designed to make them look good. Delroy L. Paulhus, "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, ed. P. R. Shaver and L. S. Wrightsman J. P. Robinson (San Diego, CA: Academic Press, 1991). Hence, performance testing will provide a more realistic picture of analytic performance if the analysts do not know they are being tested." In addition, Wilson-Wilde, Smith and Bruenisholz ("The Analysis of Australian Proficiency Test Data over a Ten-Year Period.") highlight the importance of noting "that the reasons for errors in proficiency test analysis may be different to those made in casework. Test design, differences between supplier country processes, procedures, and chemicals and test deterioration during transport may all affect the test efficacy and results obtained. Tests may also not be reflective of casework, they may be too easy (always sufficient material for testing, or a clear result is obtained), or they may be too hard (insufficient information, difficulty for suppliers to consistently produce hundreds or thousands of tests)."

[521] National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*.

[522] National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*.

Collaborative tests differ from proficiency tests in a number of ways.

1. They are not tied to meeting accreditation requirements.
2. They do not involve a registered proficiency test provider but can be created and administered by anyone (e.g., private and/or government forensic practitioners, academics).
3. They do not have to reflect casework (e.g., can focus on a portion of an examination or take a form different from real-life casework).
4. They do not necessarily have to reflect casework procedures (e.g., they could be used to validate a new test method against other methods in current practice).
5. There is no formal process for corrective action if results indicate it is needed.

Although typically based on a ground-truth-known format, collaborative trials can also be designed to test the concordance of practitioner opinion on casework material. These characteristics make collaborative trials a valuable means to investigate factors related to the claims that practitioners make. For example, collaborative trials can be used

- To validate claims or sub-claims;
- As proficiency-style tests;
- To investigate relationships between opinion profiles and experience, education, training regimes, examination times, etc.; and
- To measure laboratory, method, or FDE performance.

They can be conducted formally or informally, can test the practitioner's current skill set, and provide opportunities for skill enhancement and learning.

Perhaps the largest formal collaborative trials carried out to date were those conducted by La Trobe University in Australia from the late 1990s to the late 2000s. This institution designed and produced two trials per year (one handwriting trial and one signature trial), which yielded over 45,000 blind opinions regarding signatures and over 30,000 blind opinions on handwritten text samples.[523] FDEs from all over the world subscribed to the program, which generated valuable insights into the nature of the skills practitioners have historically claimed.

Although the La Trobe trials initially focused on the design of testing instruments that would provide data concerning claim validation and skill characterization through correct, misleading (for purposes of this report referred to as "incorrect"), and inconclusive case studies, the program quickly evolved to provide participants with data to estimate their global error rates better.

[523] See Found and Bird, "The Modular Forensic Handwriting Method—2016 Version," 63–70; Found, Sita, and Rogers, "The Development of a Program for Characterising Forensic Handwriting Examiners' Expertise: Signature Examination Pilot Study."; B. Found and D. Rogers, "The Initial Profiling Trial of a Program to Characterize Forensic Handwriting Examiners' Skill," *Journal of the American Society of Questioned Document Examiners* 6, no. 2 (2003); Found and Rogers, "Short Problem Types of Questioned Handwritten Text for Forensic Document Examiners." 8–10; Found and Rogers, "The Probative Character of Forensic Handwriting Examiners' Identification and Elimination Opinions on Questioned Signatures."

Although clients of forensic handwriting practitioners were keen to have the error rate clearly delineated, the data presented a complex and variable picture. Overall grouped scores could, however, provide some picture of the expertise (see table 4.2).

Table 4.2: Overall grouped scores for the La Trobe study questioned signature and handwriting trials

Score	Signatures[a]	Handwriting[b]
% correct	52.8	72.8
% incorrect	4.1	2.6
% inconclusive	43	24.6
% correct called[c]	92.7	96.6
% incorrect called[d]	7.3	3.4

Notes:

[a] 45,850 opinion units.

[b] 32,050 opinion units.

[c] The "% correct called" are the scores obtained after removing the inconclusive opinions and calculating the number of correct opinions divided by the total number of correct and incorrect opinions.

[d] The "% incorrect called" were calculated in an analogous way.

Variation in testing material from trial to trial, in scores among practitioners, and in the questioned writing type all affected the global scores, but the program provided two valuable opportunities.

1. Local laboratories could be informed about the profile scores of their practitioners. These scores could inform clients about the probative character of particular quality systems (or in single practitioner circumstances, the performance of that practitioner).

2. Individuals and the systems they worked within were given the opportunity to make erroneous opinions, then reflect on the opinion to revise approaches. That is, they had the opportunity to learn.

La Trobe's Revision and Corrective Action Packages (RACAP) contributed greatly to the success of the program. These results packages provided an analysis of both (de-identified) individual and group results. Participants could re-examine the images knowing what they originally opined; whether they were correct, incorrect, or inconclusive in their opinion; and knowing the responses from other practitioners.

Table 4.3 displays the opinion score profiles of a selection of FDEs (A to G) from one La Trobe University RACAP, for genuine, disguised, and simulated questioned signature types. Participants were asked to provide an opinion regarding writership on a number of questioned signatures (which were genuine, disguised, or simulated) when compared with a known signature sample set.

Table 4.3: Opinion score profiles for FDEs A to G for genuine, disguised, and simulated questioned signature types from one La Trobe University RACAP

	FDE	A	B	C	D	E	F	G
Genuine	% correct	48.3	93.3	20.8	100.0	15.0	55.0	100.0
	% incorrect	5.8	0.0	66.7	0.0	0.0	0.8	0.0
	% inconclusive	45.8	6.7	12.5	0.0	85.0	44.2	0.0
Disguised	**FDE**	**A**	**B**	**C**	**D**	**E**	**F**	**G**
	% correct	4.5	0.0	0.0	63.6	0.0	0.0	100.0
	% incorrect	18.2	0.0	90.9	0.0	0.0	86.4	0.0
	% inconclusive	77.3	100.0	9.1	36.4	100.0	13.6	0.0
Simulated	**FDE**	**A**	**B**	**C**	**D**	**E**	**F**	**G**
	% correct	79.3	15.5	20.7	100.0	0.0	87.9	46.6
	% incorrect	0.0	0.0	0.0	0.0	0.0	0.0	53.4
	% inconclusive	20.7	84.5	79.3	0.0	100.0	12.1	0.0

The table rows show percentage correct, incorrect (highlighted in red), and inconclusive opinions for each FDE, grouped by questioned signature type (i.e., genuine, disguised, or simulated). This snapshot illustrates the inter-FDE variation in score profiles across the trial's three questioned signature types. These data also provide diagnostics about practitioner cognitive strategies, or rules, that may be in use and that may be the source of incorrect/erroneous opinions.

For example, FDE D performed well on this trial, with no incorrect opinions expressed for any questioned signature type and with inconclusive opinions only recorded for the disguised category of questioned writing. Compare this result with FDE C, who expressed erroneous opinions in all but the simulated category of writing. This score profile tells us that when FDE C observed differences between the known and questioned signatures, they concluded that these were predictive of a different writer and did not fully comprehend the extent to which natural variation might be expected to occur. This latter point is why erroneous opinions were common when evaluating the genuine signatures.

Similarly, FDE F associated feature differences in the signatures with evidence of a different writer. This strategy is successful for simulated signatures (with no incorrect opinions expressed) but not for disguised signatures, evidenced by the high incorrect rate associated with disguised signatures. Meanwhile, FDE E was not confident in relation to any of the questioned signature types, opting out of expressing an opinion with regard to writership not only for all of the questioned simulated and disguised signatures but also the majority of the genuine questioned signatures.

The most important element of this collaborative program was to provide FDEs with performance metrics on ground-truth-known samples. As practitioners participated in further collaborative trials, they had the opportunity to apply lessons learned from previous trials. Trial providers hoped that the opportunity for skill improvement provided by these collaborative trials

would help mitigate error and diminish incorrect opinions in casework. Whether this occurred remains unknown. The scale of the program also provided the unique opportunity of exposure to a multitude of unnatural (disguised and simulated) writing types, which would otherwise not be available for training and development purposes.

Recommendation 4.4: The forensic document examiner community should develop collaborative testing programs aimed at monitoring and providing performance improvement opportunities related to specific claims and sub-claims. The type, content, and frequency of these collaborative tests should be determined in consultation with the research community.

4.2.6.4. Blind Declared Case

In a blind declared case, also known as a blind proficiency[524] test, the FDE (and sometimes the laboratory) is unaware that the particular case under examination is actually a test. The FDE would be aware that the workload regularly includes blind cases with a known ground truth. This type of test provides a clear indication of the performance of an examiner[525] and the laboratory system,[526] whereas a non-blind proficiency test may not.

Blind declared cases also have the advantage of countering bias because of base rate expectations, particularly for disciplines in which the FDE reaches similar conclusions for most cases. See section 2.1.5, for further discussion on base rate expectations. For example, "look-alike" non-match cases inserted into the work stream of cases for which FDEs usually make a positive identification serve to counter the base rate. This does not necessarily require double-blind testing (i.e., blind to both FDE and laboratory); a blind (to the examiner) test would suffice as long as the FDE thinks the case is real.[527] The Netherlands Forensic Institute has started a program for the inclusion of blind testing within firearms laboratories, which could serve as a model for other laboratories.[528]

4.2.6.5. Human Factors Regarding Feedback with Ground Truth Testing

Ground truth testing with timely feedback is an important aspect of building and characterizing FDE skill. As outlined, this can take a variety of forms, including black box, white box, proficiency, blind declared, competency, and collaborative tests.

Each of these tests offers laboratories and practitioners a valuable resource to test elements of handwriting evidential products that are delivered to clients; however, each has its own limitations and benefits. Generally, the tests have limited value if they assess only the expressed

---

[524] Based on the broader definition of proficiency test, rather than referring to a test required within an accreditation environment.

[525] Nikkita Venville, *A Review of Contextual Bias in Forensic Science and Its Potential Legal Implications. Report of the Victorian Law Foundation Legal Policy Internship Program*, The Victoria Law Foundation, The Australia New Zealand Police Advisory Authority and the National Institute of Forensic Science (ANZPAA-NIFS) (2011), https://netk.net.au/Psychology/Psychology14.pdf.

[526] National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*.

[527] Dror, "Practical Solutions to Cognitive and Human Factor Challenges in Forensic Science."

[528] R. D. Stoel et al., "Building the Research Culture in the Forensic Sciences: Announcement of a Double Blind Testing Program," *Science & Justice: Journal of the Forensic Science Society* 56, no. 3 (May 2016), https://doi.org/10.1016/j.scijus.2016.04.003; W. Kerkhoff et al., "Design and Results of an Exploratory Double Blind Testing Program in Firearms Examination," *Science & Justice: Journal of the Forensic Science Society* 55, no. 6 (Dec 2015), https://doi.org/10.1016/j.scijus.2015.06.007.

opinion corresponding to the known ground truth. Opinions that the examined material is insufficient or otherwise unsuitable for comparison would lead to an inconclusive opinion regarding writership (see steps 140, 210, 610, and 910 in the process map; figure 1.1). This clearly will not match the ground truth but may be entirely appropriate based on the material examined or when compared with the opinions of other suitably skilled FDEs taking the same test. This scenario was elucidated by the example of the single sans serif numeral 1, given in section 4.2.6.2.

If this limitation is acknowledged and inconclusive results are explored in the assessment of the results of ground truth tests, then they may be useful for exploring the level of agreement between opinions of different FDEs. In this way, ground truth testing can provide insight not only into overall performance but also into the concordance of FDE opinions for a particular task and help to identify errors and areas for improvement.

Other issues with ground truth testing include the problem of whether FDEs work under the same conditions and approach the task in the same way they approach case work and whether the FDEs who volunteer to participate in testing are representative of the general population of FDEs. Additionally, care must be taken to ensure that tests are designed appropriately to answer the question(s) of interest and in drawing conclusions from the results of tests. To glean meaningful findings from any data generated, a definite goal or question to be answered needs to be identified at the outset.[529]

An example highlighting these issues is the use of proficiency test data to determine error rates. CTS provides proficiency tests in various forensic science disciplines and has been asked for testing data to be used to determine error rates for specific disciplines. However, in 2010 CTS released a statement outlining why this was not appropriate.[530] The reasons included that the proficiency tests may be purchased and undertaken for a number of purposes and by a range of participants, responses are reported as in agreement or not with consensus results rather than "correct" or "incorrect," and proficiency tests are primarily designed to meet laboratory accreditation demands and may not accurately reflect casework samples.

To estimate error rates, the task itself and test samples should represent those routinely encountered in casework; using results of tests designed to be unusually difficult would be misleading. However, judicial systems might find it useful to consider different comparison types separately (e.g., comparisons of handwritten text or signatures) or samples (e.g., naturally written, disguised, and simulated) to estimate the error rate if task difficulty was comparison or sample dependent.[531]

Although not all of the material should be unusually difficult, challenging material must be included to test the limits of a system or examiner. The boundaries of FDE performance cannot be determined without pushing those boundaries until performance accuracy is affected.[532] Other

---

[529] J. B. Kadane, "Appropriate Statistics," in *Forensic Science Research Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature*, ed. E. G. Bartrick and M. A. Floyd (National Institute of Justice, 2015).

[530] Collaborative Testing Services (CTS), Inc., *CTS Statement on the Use of Proficiency Testing Data for Error Rate Determinations*, Collaborative Testing Services (CTS), Inc. (Sterling, VA, March 30 2010), https://web.archive.org/web/20170117212801/http://www.ctsforensics.com/assets/news/CTSErrorRateStatement.pdf.

[531] National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*.

[532] National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*.

issues with ground truth testing include the problem of whether FDEs work under the same conditions and approach the task in the same way as they do case work and whether the FDEs who volunteer to participate in testing are representative of the general FDE population. Additionally, care must be taken to ensure that tests are designed appropriately to answer the questions of interest and in drawing conclusions from the results of tests.

4.2.6.6. Learning Through Errors

The development of any human perceptual/cognitive skill necessarily requires feedback on the outcomes of decisions or actions.[533] This requires continual feedback about whether opinions are correct, incorrect/misleading, or inappropriate. Careful management of ground-truth-known materials, linked to specific claims to skill, is the optimal approach for acquiring the necessary skills to attain competency for the cognitive task.

Most training in forensic handwriting follows the mentored or apprenticeship approaches. In these modes, trainees carry out much of the casework under the supervision of a suitably qualified mentor. In many parts of the world, handwriting examination is only one of several competencies required of the trainee. Others include examinations of print processes, indentations, alterations, obliterations, and erasures. The training period usually ranges from 2 to 5 years but can be longer. Although mentored training has been the accepted approach, very little information exists about the standards and metrics mentors employ to evaluate competency throughout training processes. In addition, training programs that focus on casework depend entirely on the mentor's skill, and the ground truth is usually not known in casework. Furthermore, the extent to which competency in handwriting is assessed by mentors using casework samples compared with an independent assessment using ground-truth-known samples remains largely unreported.

Claims to expertise should be linked to standardized and validated ground-truth-known collaborative testing materials that represent the various tasks and difficulty levels encountered in casework. These collaborative tests should not only be aimed at addressing holistic tasks (which one might expect to look like casework) but would also focus on the many sub-tasks that contribute to higher-level decision-making activities.[534]

Recommendation 4.5: The forensic document examiner community should develop a framework for feedback-driven training, testing, and development based on ground-truth-known material.

[533] K. A. Ericsson, R. T. Krampe, and C. Tesch-Romer, "The Role of Deliberate Practice in the Acquisition of Expert Performance," *Psychological Review* 100, no. 3 (1993); K. A. Ericsson, "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance," in *The Cambridge Handbook of Expertise and Expert Performance*, ed. N. Charness K. A. Ericsson, P. J. Feltovich, and R. R. Hoffman (Cambridge University Press, 2006), 685–706; K. A. Ericsson, *Peak: Secrets from the New Science of Expertise* (Boston: Houghton Mifflin Harcourt, 2016).

[534] Ericsson, "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance."

4.2.6.7. Tracking the Outcome of a Forensic Analysis: Beyond Simple Errors

In impression and pattern evidence disciplines, two outcomes of an analysis are often characterized as matches or non-matches.[535] An FDE's performance is then characterized by examining the number of correct and incorrect responses relative to ground truth, which may be further split into false positives and false negatives. The statistical tools to describe this type of binary response model are well-developed and widely used. The concepts of sensitivity and specificity of forensic test procedures are based on this description of the outcomes, as limiting as the descriptions may be. However, FDEs currently use a multi-point scale, typically with three to nine outcomes of varying weight of evidence (see chapter 1, table 1.4).

Therefore, any model of error regardless of the point scale used should account for opinions by the FDE that the evidence is either insufficient (see steps 140, 210, 610, and 910 in the process map; figure 1.1) or inconclusive (see step 1320 in the process map; figure 1.1). These categories, if not considered, may skew the results of a proficiency test by suggesting that FDEs who are excessively conservative in their opinions are less proficient than those who are less conservative. That is, in an environment where inconclusive/insufficient responses are not tracked and FDE responses are "marked" against the ground truth, a more conservative FDE may be considered less proficient because they will not give a response that is the same as the ground truth (and therefore they will be marked "wrong"), whereas a less conservative FDE may give the "right"/ground truth answer. The conservative response, however, may be the most appropriate response.

Whether inconclusive opinions should be considered incorrect is a matter of debate among FDEs, researchers, and legal professionals. For instance, one may argue that inconclusive opinions are correct opinions intended to indicate that the writing samples are insufficient for comparison purposes, regardless of whether ground truth is known. Others may argue that the excessive use of an inconclusive finding may be inappropriate and overly cautious. Studies show that error rates for handwriting examination tend to be significantly higher when inconclusive opinions are counted as errors.[536] Studies have also shown that skilled FDEs are more effective than the general populace in determining when the evidence is insufficient to make a decision.[537]

With respect to fingerprint examinations, the Latent Print report[538] presented an argument against the inclusion of "insufficient" or "inconclusive" in the calculation of error rates, as stated by Koehler:[539]

---

[535] M. Houck and J. A. Siegel, *Fundamentals of Forensic Science* (Burlington, MA: Academic Press, 2009), 281. Also firearms "match" in J. Song, "Proposed "NIST Ballistics Identification System (NBIS)" Based on 3D Topography Measurements on Correlation Cells," *AFTE Journal* 45, no. 2 (2013): 184–94, https://pdfs.semanticscholar.org/1103/38f65f80958b94317bfd780556231adf0995.pdf.

[536] B. Found, D. Rogers, and A. Herkt, "The Skill of a Group of Document Examiners in Expressing Handwriting and Signature Authorship and Production Process Opinions," *Journal of Forensic Document Examination* 14 (2001).

[537] Bird, Found, and Rogers, "Forensic Document Examiners' Skill in Distinguishing Between Natural and Disguised Handwriting Behaviors," 1292–94; Found, Sita, and Rogers, "The Development of a Program for Characterising Forensic Handwriting Examiners' Expertise: Signature Examination Pilot Study."; Kam et al., "Signature Authentication by Forensic Document Examiners."; Sita, Found, and Rogers, "Forensic Handwriting Examiners' Expertise for Signature Comparison."

[538] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 29.

[539] J. Koehler, "Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter," *Hastings Law Journal* 59, no. 5 (2008): 1080–81.

> When an examiner offers an "inconclusive" opinion about whether two prints match, there is a sense in which he has erred. After all, he did not get the answer right, and the consequences of this failure may be serious (e.g., missed opportunity to exonerate a suspect). However, in the more usual sense of the meaning of error, an inconclusive is not an error. It is a pass. An inconclusive means that the examiner offers no judgment about whether two prints do or do not share a common source.

In contrast to this viewpoint, the Bromwich report[540] cited an inappropriate application of the inconclusive category:

> Derrick Leon Jackson is a death row inmate who was convicted in a capital murder case in which the Crime Laboratory performed extensive serological testing. In 1988, Mr. Bolding obtained ABO typing results from a bloodstain sample taken from the scene of a grisly double homicide that indicated the sample was foreign to both the victims and the individual whom investigators originally suspected of the killings. At the time, however, Mr. Bolding reported these results as "inconclusive," perhaps because the results were not consistent with investigators' initial theory about who may have committed the crime. The investigation languished until 1995 when Mr. Jackson became the prime suspect. Mr. Jackson's ABO type was consistent with the foreign ABO factor Mr. Bolding had detected in 1998, which he originally described as "inconclusive." Without performing any additional testing, Mr. Bolding altered his worksheets to include previously absent conclusive interpretations of his original typing results performed in 1988 and issued a new report stating that ABO activity consistent with Mr. Jackson's ABO type was found in two bloodstain samples recovered from the crime scene.

The process map included in this report (figure 1.1) combines the two categories of insufficient and inconclusive into a single outcome (step 1320), fed into from various steps in different stages of the process map (see, for example, steps 170–200 in the pre-analysis stage and step 1180 in the evaluation stage). In practice, the Working Group recognizes that protocol in at least some laboratories will require that the reasons for the inconclusive/no opinion conclusion is documented and reported. For QC purposes, it would be preferable to track the insufficient and inconclusive categories separately. Tracking these forensic analysis outcomes makes it easier to document the performance of a laboratory (via proficiency tests or casework product) or individual FDEs. If the insufficient category is invoked at widely different rates between FDEs or between laboratories, it might indicate an area where improvements could be made. To date, researchers have not conducted enough ground truth studies to determine empirically supported best practices in this area.

Overstating or understating the meaning of evidence has caused severe problems in forensic science.[541] If the level of certainty or quality of evidence is exaggerated, this is a flawed outcome. Although the results of an examination may be correct (matching ground truth), reporting the results, either written or verbal in courtroom testimony, may overstate or understate

---

[540] M.R. Bromwich, *Final Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room*, Fried, Frank, Harris, Shriver & Jacobson LLP (Washington, DC, 2007), 95, 96, http://www.hpdlabinvestigation.org/reports/070613report.pdf.

[541] This stance is taken in an FBI press release regarding its "Microscopic Hair Comparison Analysis Review." Federal Bureau of Investigation, "FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review," news release, 2015, https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review.

the weight of the evidence or the level of certainty in the conclusion. Result tracking, both in case work and in testing situations, needs to incorporate some method to detect and record understatements and overstatements of the certainty of results. For example, CTS proficiency tests allow the test taker to state that the samples "cannot be identified or eliminated." However, the FDE does not have the opportunity to conclude that the samples were deemed insufficient to make a determination.[542]

Recommendation 4.6: Quality control procedures should include tracking of inconclusive and insufficient opinions. Test materials should include these opinion categories.

4.2.7. Documentation and Record Keeping

Documentation is a multi-faceted component of any QMS. The QMS must clearly define policies, procedures, organizational outlines, and management duties. Management system documents should be authoritative, periodically reviewed, and properly maintained. These documents may include general laboratory and safety policies, evidence bulletins, test methods, and training programs.

Documentation is also essential to describe the improvements made to the organization or the individual through competency and proficiency testing, continuing education, implementation and validation of procedures, audits, and the results of any corrective actions to resolve significant technical problems. A policy should be in place to track and control revisions and periodic updates to QMS documents. This ensures that the most up-to-date procedures are applied and referenced both internally and externally, while also providing a record of any changes made within the system.

Documentation must be contemporaneous regarding the handling and continuity of the evidence, the procedures used within the case examination, and the monitoring of the quality of the work through case review and courtroom assessment. Recording the evidence, activities, and results at the time they are acquired or occurred aids review, testimony, research, and improvement activities. The documentation should lead an independent FDE to understand the process of continuity and evidence handling, the methods used within the examination process, the basis of any opinion formed, the relationship between the opinion and the reporting scale, and any limitations of the examination method. Additionally, explicit documentation of the bases for an opinion greatly aids the interpretation and review process.[543] The QMS should provide clear guidance on what information should be included in both the case notes and report. Report writing is covered extensively in chapter 3.

The extent of documentation in the case record may vary according to the FDE's assessment of case complexity, feature selection, and sufficiency of the evidence for examination. Without national minimum standards for documentation and report writing, QMS requirements may vary

[542] Collaborative Testing Services (CTS), Inc., *Handwriting Examination Test No. 17-523/524 Summary Report* (2018), https://cts-forensics.com/reports/3724_Web.pdf.

[543] The Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, 41.

between laboratories, which could make case and testimony review across laboratories challenging.

4.2.8. Personnel, Accommodation, and Environmental Conditions

At a minimum, the laboratory should contain adequate space for equipment and employees, secure areas for evidence storage and handling, and a health and safety program for employees. The QMS should maintain the records and provide oversight for training, certification, and testing for the personnel. The quality and management personnel should work together to define satisfactory completion of testing and identify the appropriate actions to take when employees fail to achieve the expected results. Chapter 5 reviews training, whereas chapter 6 covers in more detail some personnel qualifications and environmental and accommodation conditions.

This page intentionally left blank.

5. Education, Training, and Certification

Introduction and Scope

Proper education and training are the building blocks on which an FDE gains and maintains expertise; appropriate education and training also minimize human error in the examination process. This chapter reviews the education and training an FDE must master. Foundational education refers to the academic prerequisites that qualify an individual for forensic handwriting examination training. The specialized training that follows focuses on the discipline-specific requirements and competencies necessary for an individual to qualify as an FDE. This chapter also addresses how certification[544] can tie many of these related issues together. Once deemed competent, the FDE maintains currency in the discipline by continuing their education. Given that communication is such a critical human factors issue, training should focus on teaching the best way to convey information to investigators and triers of fact in an attempt to minimize errors associated with miscommunication.

5.1. Foundational Education

Adequate foundational education, coupled with testing, provides the core competencies on which proper training can be built. The Working Group identified several core competencies, each of which provides an appropriate educational foundation and skill set and should be demonstrated by training candidates. The core competencies most related to the FDE role include

- Science, technology, engineering, and mathematics (STEM);
- Psychology (cognitive skills, social sciences, and form blindness testing);
- Probability and statistics;
- Literacy skills (including the ability to read and write cursive, reading comprehension, active listening, clear oral and written communication skills, and technical writing skills);
- Computer skills;
- Critical thinking;
- Physiological capabilities (including corrected eyesight, attention, and concentration); and
- Research methodology.

Government laboratories typically require a college degree for employment, which will generally require the completion of courses that encompass the listed topic areas. Although many highly qualified FDEs do not have college degrees, the Working Group concluded that a college degree and accompanying transcripts provide the best avenue for verifying completion of the prerequisite academic-related core competencies. In addition to opening more opportunities for

[544] Certification is not the same as accreditation. Certification assesses an individual's competence, whereas accreditation only assesses the laboratory as a system. U.S. Department of Justice, Office of Justice Programs, *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*, Technical Working Group for Education and Training in Forensic Science (TWGED) (April 2004).

employment, several professional organizations, including the AAFS,[545] require a college degree for membership. Finally, FDEs who do not possess such a degree may find that their analyses are considered with less weight.[546] However, the Working Group recognizes that college or university degrees are not the only method of obtaining the required level of knowledge in the core curriculum. Those who have chosen alternative routes like individualized course work, work experience, and training courses will need to provide ample documentation of their ability to satisfy these competencies like coursework syllabi, training agendas and materials, a resume or CV, or authored publications.

Some of the core capabilities are not academic. These include eyesight, the ability to differentiate patterns, oral communication, and the ability to concentrate. These capabilities should be tested in each candidate. Candidates who have physiological limitations like form blindness and color blindness may not be capable of performing forensic handwriting examinations.

5.2. Training

The current methods of training in the United States vary greatly (including self-taught and apprenticeship models among others) and therefore may not always allow for a uniform program or a consistent and rigorous evaluation of an individual's training progress and competence. For example, Behrendt wrote in 1989 of the many difficulties encountered in training FDEs, many of which are still relevant today:

> Questioned document examination has traditionally used on-the-job training as its primary instructional method. There are several deficiencies inherent in this method of training, however. Some of these deficiencies are the lack of a standardized course of instruction, the inability to evaluate the quality of the training received by an individual, the absence of any criteria establishing minimum levels of competency, and the length of time required which results in a reluctance to hire trainees.[547]

Forensic document examination encompasses several forensic disciplines (examinations of handwriting, typewriting, printing processes, indented impressions, alterations, and ink, as well as advanced processes like Fourier Transform Infrared and Raman spectroscopy), each requiring different skills and examination techniques. Requiring that an FDE must achieve knowledge, skills, and abilities in all areas of questioned documents to be deemed competent may be a dated notion and leaves many of the challenges encountered in both the public and private sectors unaddressed.

Globally, training approaches and competence vary. Some organizations take a holistic approach, requiring that individuals be trained in every possible aspect of their chosen field of work. In

---

[545] "Types of Forensic Scientists: Disciplines of AAFS," 2019, accessed May 6, 2020, https://www.aafs.org/aafs/Resources/Students/Choosing-a-Career/Types-of-Forensic-Scientist--Disciplines-of-AAFS/AAFS/Resources/Students/Types.aspx?hkey=a401388a-32dc-4c74-aee5-c52de769c6e0.

[546] M. L. Merlino, C. I. Murray, and J. T. Richardson, "Judicial Gatekeeping and the Social Construction of the Admissibility of Expert Testimony," *Behavioral Sciences and the Law* 26, no. 2 (2008), https://doi.org/10.1002/bsl.806; M. L. Merlino et al., "Meeting the Challenges of the Daubert Trilogy: Refining and Redefining the Reliability of Forensic Evidence," *Tulsa Law Review* 43, no. 2 (2008), https://heinonline.org/HOL/LandingPage?handle=hein.journals/tlj43&div=25&id=&page=; M. L. Merlino, V. Springer, and A. Sigillo, "The Social Construction of the Admissibility of Most Frequently-Proffered Varieties of Expert Testimony," in *The Future of Evidence: How Science and Technology will Change the Practice of Law*, ed. C. Henderson and J. Epstein (Chicago: American Bar Association, 2011), 1–20.

[547] James E. Behrendt, "The Status of Training for Questioned Document Examiners in the United States," *Journal of Forensic Sciences* 34, no. 2 (1989), https://doi.org/10.1520/jfs12645j.

contrast, other organizations employ a discipline-specific approach. Someone specializing in handwriting examination need not be an expert in all areas of document examination but must have adequate knowledge of other aspects like alterations, print processes, and indentations so that they can best preserve the evidence and alert other specialists to potential evidence that may require additional examination. Similarly, an expert in electrostatic detection of indented impressions on documents does not necessarily have to be an expert in handwriting comparisons but must have sufficient knowledge to appreciate the potential forensic value of various observations. Training and competence for each specialization should be transparent and consistent.

Routinely, FDEs are trained through apprenticeship with an expert helping to lay down a foundation of knowledge and experience through instruction and explanation of laboratory protocols. However, this individualized apprenticeship approach alone may not always be the most effective mechanism for training an FDE,[548] as discussed in section 4.2.6.8.

5.2.1. History of Training Standards

In 1942, the first professional FDE organization was incorporated. This organization consisted of FDEs in the private sector who had met regularly but informally for over 30 years, often at the home of Albert S. Osborn.[549] One agenda item established that membership would require applicants to have completed 3 years of training. This requirement was later modified to 2 years. FDEs from the public sector were subsequently admitted to the organization under the same training requirements.

In 1977, the first certification body was established with funding from a Law Enforcement Assistance Administration grant and sponsorship/recognition by two significant forensic research bodies. From its inception, this certification board required each applicant to have completed a minimum of 2 years of training. Numerous other forensic document examination professional organizations formed over the past 40 years have required the same amount of training.[550] As such, this length of training has long been accepted within the United States for experts in both the public and private sectors and has been a requirement for applicants for positions at numerous law enforcement crime laboratories. A minimum of 2 years of training has been a requirement of most public sector laboratories for at least 50 years. The booklet *Objectives for Training*[551] noted the requirement of 2 years of training. It also noted that any specialized training that might result from an individual's particular employment would be in addition to the 2 years of basic training.

In 2005, the discipline established a codified *Standard Guide for Minimum Training Requirements for Forensic Document Examiners* (published by ASTM[552]), setting a minimum of

[548] Behrendt, "The Status of Training for Questioned Document Examiners in the United States."

[549] Albert S. Osborn is considered the "father of forensic document examination," having published the seminal textbook *Questioned Documents* in 1910.

[550] Behrendt, "The Status of Training for Questioned Document Examiners in the United States."

[551] American Board of Forensic Document Examiners (ABFDE), *Objectives for Training*, 2 ed. (2010).

[552] *Standard Guide for Minimum Training Requirements for Forensic Document Examiners,* ASTM E2388-11, (West Conshohocken, PA: ASTM International, 2011).

24 months of training within a 4-year period or equivalent. In 2012, SWGDOC adopted the ASTM training standard and currently maintains that standard.

The term "equivalent" has been used in conjunction with the length and format of training in published minimum standards for training. The Working Group has seen a trend toward misapplication of this term. Equivalent is frequently used to denote different ways that one may obtain proper training of over 4,000 hours within 4 years. However, equivalency cannot be achieved solely by distance learning, periodic phone conversations, or even periodic face-to-face meetings. Although some aspects of forensic document examination (e.g., court procedures, evidence handling, scientific method, historical foundations, research methods, print process, paper and ink identification methods, and copybook styles) may be effectively taught through various formats, the intricacies of handwriting and signature identification are not conducive to online or distance training. Although there are many activities necessary to building competencies in forensic document examination, training in handwriting and signature examinations requires detailed, in-person, one-on-one instruction between trainer and trainee and should constitute the majority of the training program.

Explaining and demonstrating the subtleties of handwriting execution, natural variation, and fundamental differences is best achieved through in-person instruction with immediate feedback. Studies conducted on the efficacy of online distance education programs support the contention that some disciplines (e.g., chemistry laboratory, biology laboratory, physics laboratory, osteology, dental hygiene, health sciences laboratory, skilled labor fields) require "brick and mortar" avenues for effective learning.[553] Just as one may not wish to be treated by a physician trained solely through online instruction, the same may be said of an FDE testifying in a case in which an individual's liberty hangs in the balance.

As shown in table 5.1, a 2014 study[554] of 97 U.S. FDEs found the average length of formal training to be 2.5 years with a range of 1 to 6 years.

Table 5.1: Information relating to length of training and experience of FDE

Forensic Document Examination Training	Minimum	Maximum	Average (Mean)	Standard Deviation
Length of FDE training (years)	1	6	2.5	.79
Since FDE training completed (years)	0	42	19.9	11.5

On average, FDEs completed their training approximately 20 years ago. In Europe, the training program for a forensic handwriting expert varied from 6 months to 5 years (n = 216),[555] depending on the qualifications of the individual and specific organizations' requirements.

[553] E. Verma, "From Traditional Classroom to Online Learning: ICT as Change Agent to Boost Innovation in Teaching and Learning," *International Journal of Scientific Research in Science and Technology* 3, no. 8 (2017). See also, "The International Review of Research in Open and Distributed Learning home page," 2019, accessed May 6, 2020, https://www.irrodl.org.

[554] Merlino et al., *Validity, Reliability, Accuracy, and Bias in Forensic Signature Identification*.

[555] Internal study undertaken within ENFHEX on training processes.

5.2.2. Training Manuals

The numerous laboratories that train FDEs have a variety of training manuals. The U.S. Army Crime Laboratory has had a training manual[556] for forensic document examination since the 1960s, as have other federal laboratories and state law enforcement agencies. One Working Group member examined several manuals (with the understanding that the manuals would not be distributed) and found that they had highly similar training outlines. However, the reference papers on which the manuals were based were weighted heavily toward experts working in the same geographic region as the publisher of the manual. The designated time frame for each section of training varied greatly. The OSAC is developing a standard training program by subject based on current methods of training within the United States.[557] Training of competent FDEs in the public sector generally follows the proposed "Standard Training Program for Forensic Document Examiners."[558] However, the Working Group identified three issues that need to be addressed:

1. The specification that the training must be for at least 24 months.
2. The notion that training must be at least 4,480 hours (this equates to 320 days per year at 7 hours per day), which the Working Group believes is not realistic. The actual amount of training time, depending on the modules completed, among other variables, may take less or considerably more time.
3. The competence process is designed as "pass a competency test," but no details are given as to how that process should be evaluated.

5.2.3. Current Training Processes

Based on the U.S. training manuals reviewed by the Working Group, a subject-by-subject method of training appears to be the standard and is generally accepted within the United States as the best practice. Historically, trainees were (1) trained under the tutelage of FDEs either in private practice or in government laboratories in an apprenticeship or mentorship capacity, (2) tested by the trainer, and then (3) certified by a body of FDEs. In Europe, whether the training is designed to create an expert covering all aspects of FDE or specific areas, the training is carried out in a modular format. The ENFSI[559] published a template and proposal for forensic handwriting examination training in a best practice manual now being adopted across Europe. Furthermore, the National Institute of Forensic Science, a directorate within the Australian and

[556] U.S. Army Criminal Investigation Laboratory, *Program of Instruction for Document Examination Course* (October 1966).

[557] OSAC was considering ASTM E2388-11 (ASTM E2388-11, 2011) as an OSAC standard and released the standard for a Public Comment Period, which has closed. This standard has been withdrawn from the Standards and Public Comment Adjudication Phase at the request of the Forensic Document Examination Subcommittee until further action is taken by the Subcommittee. Organization of Scientific Area Committees (OSAC) for Forensic Science, *OSAC Standards Bulletin* (2017), https://www.nist.gov/system/files/documents/2017/10/24/osac_standards_bulletin_october_2017_2.pdf.

[558] ASTM International. *Standard Guide for Minimum Training Requirements for Forensic Document Examiners*.

[559] European Network of Forensic Science Institutes (ENFSI). *Best Practice Manual for the Forensic Examination of Handwriting*.

New Zealand Policing Advisory Agency, developed *Guidelines for Education and Training for Forensic Document Examiners*.[560]

The European system takes the trainee through each facet of the relevant examination topic by topic, allowing the trainee to absorb the information in an orderly form. Each module includes four parts:

1. Laboratory protocol (evidence handling, evidence protection, evidence marking, chain of custody);
2. Instruction (providing the fundamental and foundational learning of the subject, including reading texts and papers; attending lectures; training in instrumentation, methodology, statistical implications, report writing, and testimony; and examining mock cases [e.g., with ground truth results]);
3. Experience foundation (multiple cases of a diverse range); and
4. Assessment (continual accuracy in casework and successful completion of tests as basis for advancement to next step).

There are two principal differences (although others do exist) between U.S. and European approaches to training.

1. Training in Europe and other countries is moving toward a competence assessment approach in contrast to the conventional U.S. system of having a minimum time for training before testing competence. A proposed European personal certification process for forensic scientists also addresses FDE training.
2. Unlike the U.S. method of general qualification, training in Europe separately qualifies handwriting experts, document experts, ink specialists, and document and handwriting experts. The training processes for these disciplines are modular, and an expert can be deemed competent in one area without having to be deemed competent in another.

Forensic handwriting examinations generally constitute the bulk of examinations conducted by an FDE. Some FDEs specialize in handwriting and consult with other specialists in the fields of document examination when it appears they may be needed. In addition to handwriting identification, many certified FDEs in the United States conduct forensic examinations in related specialized fields like electrostatic latent imaging (e.g., electrostatic detection device), ink analysis (thin layer chromatography, Fourier, Raman, etc.), alterations made to questioned documents, and print process identification. Often, the FDE is asked to authenticate a document on which a signature may appear. Although the signature may be authentic, the FDE must also consider the possibility that the signature was cut and pasted onto a document or that pages or printed material may have been inserted into the document. This requires at least a working knowledge of fields related to handwriting identification (e.g., electrostatic detection device, print process identification, ink and paper examinations, computer-generated documents). As

[560] "Education and Career Information," n.d., accessed May 6, 2020, https://www.anzpaa.org.au/forensic-science/forensic-sciences/education-and-career-information.

such, training modules in forensic document examination, even for those focused on handwriting, should include these areas so FDEs will know who and when to consult if that area falls outside the realm of their expertise.

5.2.3.1. Current State of Education and Training

Formal education opportunities in forensic document examination are limited. For instance, the Working Group identified among 126 U.S. tertiary institutions 203 degree-level forensic science programs ranging from certificates to doctoral degrees. About half of the programs were at the bachelor's level. At the time this report went to press, the Working Group had identified only three programs providing more than just a one-time overview of forensic document examination.[561]

University courses provide an unparalleled opportunity to expose students to the field of forensic document examination, but these programs appear to be limited in number. Moreover, the Working Group has become aware of numerous candidates with advanced degrees and passionate interest in the discipline who are unable to obtain proper training because of limited resources for training, testing, and career development.

5.2.3.2. A Future Vision for Education and Training

The Working Group concluded that the lack of formal training opportunities is the largest obstacle to recruiting new people to the field and producing properly trained FDEs in both public and private sectors.

The first step in correcting this limitation is identifying organizations with adequate resources to house and administer training in forensic document examination on a regular basis and that are open to public and private sector students. Universities and centers of excellence are examples of the types of organizations that may be suited for these endeavors.

The second step is establishing an overall project plan, which should include the following:

- A comprehensive list of necessary start-up equipment, personnel, and support;
- Establishment of an acceptable training program to include all necessary training equipment and other training material, available supplemental workshops, and consulting instructors;
- An avenue to conduct the significant amount of foundational research that this report is advocating; and
- A list of student grant, loan, and scholarship sources to assist those who apply for training.

This vision is undoubtedly a major and expensive undertaking. However, the Working Group offers the following examples as potential ways to mitigate the financial burden.

[561] The Working Group identified certificate programs at East Tennessee State University and University of Baltimore, and a Forensic Document Examination track for a master's degree in Forensic Science at Oklahoma State University.

- Several universities house and administer funded research. Funded projects normally include a percentage designated for administration. As such, it is anticipated that certain universities would find this proposal inviting.
- As part of establishing a research and training laboratory, the laboratory would accept contract casework for investigative, prosecutorial, and defense entities. This casework would generate funds for the laboratory to offset costs and real casework for the development of core experience by the trainees.
- Manufacturers of specialized equipment need field testing; a research and training laboratory would be an ideal source for new product testing and evaluation. By partnering with equipment manufacturers, the research and training laboratory may garner favorable considerations when purchasing equipment.
- Students will attend classes for credit as an integral part of training. As such, the student will obtain advanced degrees commensurate to the time and effort for training and the laboratory/university team will be able to offset expenses by the tuition fees charged. As an added benefit, this plan will produce trained FDEs with advanced degrees.
- The laboratory subject matter experts will also serve as faculty members for classes that include paying students. Additional undergraduate classes could also be taught by these experts.

To support this vision further, the Working Group suggests that the federal government provide funding in the form of a grant to establish a forensic document examination research and training laboratory open to both public and private sector students.

Recommendation 5.1: To improve training, forensic document examiner professional organizations and practitioners should pursue both private and government funding like scholarships, grants, or loans to offset training costs.

5.2.4. Cross-Training

Many agencies are downsizing or eliminating departments with expertise in handwriting examination.[562] Furthermore, the FDE population is aging; on average, active FDEs have been in the field for more than 20 years (see table 5.1). The danger looms that as the number of experienced FDEs dwindles, there may not be enough experts to train and mentor the next generation. One way that full-service forensic laboratories can help maintain or increase the number of trained FDEs—without adding to the total number of staff—is by cross-training forensic specialists in more than one discipline.

For example, at the Los Angeles Sheriff's Department Crime Laboratory, plans are underway to cross-train FDEs so that they can perform analyses in other forensic areas like shoe and tire impressions or gunshot residue. This type of creative management can help to ensure the longevity of the discipline. A similar process already exists in the Chemistry and Documents

[562] See Table 1 of Durose et al. *Publicly Funded Forensic Crime Laboratories: Quality Assurance Practices, 2014*. The percentage of laboratories reporting on questioned documents is decreasing: 24% (2002), 20% (2005), 16% (2009), and 14% (2014).

Team of the Scottish Police Authority, Forensic Services in Scotland and in the Chemistry Section at Forensic Science SA in Australia.

5.2.5. Trainers

The SWGDOC minimum training standard[563] requires that trainers be certified FDEs who have undergone training that meets published standards. Trainers should have also achieved recognition as educators (through an appropriate degree, documented classroom and educational experience, or attendance at trainer-skill workshops). Trainers are expected not only to possess the knowledge, skills, and abilities of a certified FDE but also to be able to impart those traits to a trainee. Additionally, trainers are expected to develop general lesson plans, learning objectives, learning outcomes, course syllabi, and testing and evaluation methods for trainees (if these are not already part of the laboratory's training manual), and document training activities and trainee transcripts.[564]

Trainers should receive formal training in instructional skills through college-level courses or workshops facilitated by professional societies. Trainers in accredited laboratories may have their own specific requirements for training officer qualifications.

Recommendation 5.2: Academia and professional forensic document examiner organizations should collaborate to develop trainer-skill workshops and classes.

5.2.6. Future of Training for FDEs

A forensic document examination may consist of more than just "handwriting examinations." A modular approach to training can offer support for other examination areas without the need to be competent in all of them. Different people in different organizations require different skill sets, and the FDE community should develop a process that allows for this. To challenge the need for time-specific constraints in training, forensic handwriting training must employ robust learning methodology, freely borrowing from academia (in the form of a revised Bloom's Taxonomy[565] [see figure 5.1]) various ways to approach training and development.

An academic, modular process should be adopted by the forensic document examination community to develop the highest quality practitioners working within the field, as noted in Recommendation 5.3. In general terms, the process would be based on a tiered system of training, each tier providing ever-increasing knowledge, skills, and abilities to the trainee, culminating in a final set of competency tests managed and overseen by a body or panel independent of the FDE's workplace.

A fixed time scale may not be the best method for training FDEs. People learn and develop at different rates and any training, and despite maintaining consistency in the curriculum and materials, the scale should be adjusted timewise to individual requirements. By developing

---

[563] Scientific Working Group for Forensic Document Examination (SWGDOC). *SWGDOC Standard for Minimum Training Requirements for Forensic Document Examiners*. Section 5.5.

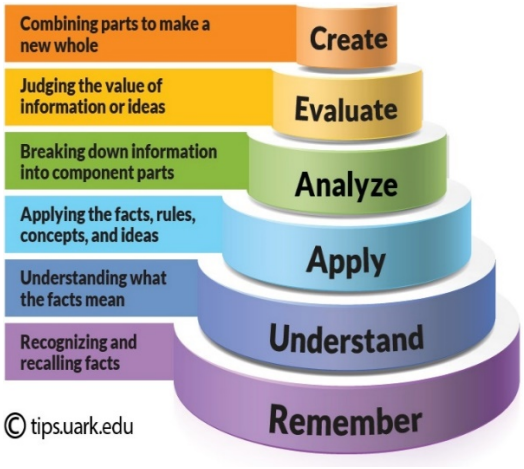
[564] For example, the University of Kentucky offers a Preparation of Future Faculty program that specifically addresses teaching pedagogy. "About PFCF/PFF," University of Kentucky, n.d., accessed May 11, 2020, https://www.uky.edu/CommInfoStudies/GRAD/PFF/about.html.

[565] B. S. Bloom et al., *Handbook I: Cognitive Domain*, Taxonomy of Educational Objectives: The Classification of Educational Goals, (London, WI: Longmans, Green and Co LTD, 1956); Peter W. Airasian et al., *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, ed. L. Anderson and D.R. Krathwohl (Boston: Allyn and Bacon, 2001).

specific learning outcomes allied to the elements of the cognitive domain section of the revised taxonomy, a more robust and individually focused training program can be developed. However, some may erroneously claim that training over a few short weeks or months is adequate. To address this, FDEs need to successfully complete a robust competence test for each of the training modules contemporaneous to their development.

Training is divided into a number of key stages (e.g., introduction, foundation, reinforcement, consolidation, and reporting). For each stage, the various modules undertaken by the trainee will have a series of defined outcomes. Two possible elements in the proposed training program are provided in table 5.2, which outlines a knowledge component in the foundation stage, and table 5.3, which outlines a practical component in the reporting stage.

Bloom's Revised Taxonomy, published in 1956, is a classification system designed to improve communication between educators and students and to establish more suitable curricula for education. Consisting of three domains—knowledge-based, emotive-based, and action-based (also referred to as the cognitive domain, the affective domain, and the psychomotor domain, respectively)—each domain was divided into various descriptive "learning" objectives. In 2001, the cognitive domain was revised by Anderson and Krathwohl to convert the text to a more "active" prose (see figure.) Anderson and Krathwohl described the elements of "remembering" and "understanding" as being "lower order thinking skills," whereas "evaluating" and "creating" are considered to be "higher order thinking skills."



The concepts within this process allow for a rigorous and structured approach to education and learning, applicable to a wide range of topics.

Figure 5.1: Bloom's Revised Taxonomy

Table 5.2: Hypothetical knowledge component of a foundation stage topic in a proposed training program

Module	Handwriting examination and comparison (including signatures)—general.
Module Objective(s)	The purpose of this module is to introduce the trainees to the types of handwriting routinely encountered.
Trainee Learning Objective(s)	Trainees will be able to define the differences in natural, disguised, traced, and simulated handwriting. Trainees will be able to describe the characteristics of each type of writing. Trainees will be able to discuss the differences between natural, disguised, traced, and simulated handwriting.
Assessment Method(s)	Trainees' ability to define differences in handwriting will be measured by undertaking a multiple-choice questionnaire covering the various types of handwriting encountered. Trainees' ability to describe the differences between the types of handwriting will be measured by written essays and oral presentation of information.
Success Benchmark(s)	Successful completion of this module will be demonstrated by a correct response rate of at least 95% in the multiple-choice questions and a mark of at least 85% in the written essay and oral questioning.

Table 5.3: Hypothetical practical component of a reporting stage topic in a proposed training program

Module	Handwriting examination and comparison (including signatures).
Module Objective(s)	The purpose of this module is to test the trainees on their ability to report a large, complex handwriting examination.
Trainee Learning Objective(s)	The trainees will be able to demonstrate the procedures involved in a large handwriting examination.
Assessment Method(s)	Trainees' ability to demonstrate the handwriting comparison process will be measured by undertaking a number of complex ground-truth-known handwriting comparisons covering the various types of handwriting encountered. Each of these comparisons and their outcomes will be assessed by an independent verifier, for example the trainer or another peer.
Success Benchmark(s)	Successful completion of this module will be demonstrated through an assessment by the independent verifier reviewing both the case notes and the final reports. The assessment will include an oral questioning component. Success will be contingent on at least 90% achievement for all three aspects of the assessment (case notes, report, and oral questioning).

5.2.6.1. Introduction Tier

The training considers that many of the fundamentals in forensic science are not discipline-specific and can be covered in a generic process. In the suggested training program, an Introduction tier covers these fundamentals under modules like those listed below.

- Introduction to forensic science
- Introduction to quality management
- Crime scene preservation
- Evidence handling
- Note taking
- Introduction to ACE-V process
- Statement and report writing
- Criminal justice systems
- Training in the cognitive aspects of forensic science (including the effects of bias)
- Statistics, probability, and interpretation of findings

- Literature—particularly pertaining to forensic handwriting examination

Each module, based on Bloom's Revised Taxonomy, is associated with a series of specific module objectives, learning objectives, assessment methods, and success benchmarks (as illustrated in table 5.2 for a component of the foundation tier and in table 5.3 for a component of the reporting tier). At the end of the introductory training period, the trainee will undertake a series of competence assessments relating to the above skills.

5.2.6.2. Foundation Tier

After completing the Introduction tier, the trainees move to the Foundation tier. In this tier, the trainees become acquainted with the fundamentals of the forensic science area in which they will be trained and eventually reach full competence. Modules covered in this level include general areas like examinations of documents for fingerprints and DNA and counter-contamination protocols but also the foundation levels of questioned document examination, including components both related and not related to handwriting. Areas covered include the fundamental principles of

- Indented impressions examinations (including electrostatic detection device and oblique light),
- Handwriting examination and comparison (including signatures),
- Altered documents,
- Conventional printing examinations,
- Office printing systems and output,
- Paper examinations,
- Dating documents,
- Chemical ink analysis, and
- Digital writing and related issues.

Similar principles to those used for competence assessment in the Introduction tier will be employed and cover specific module objectives, learning objectives, assessment methods, and success benchmarks.

At the culmination of this tier, the trainees progress to examination-specific modules for the Reinforcement and Consolidation tiers. An agreement between each trainee and their trainer—and where relevant, in accordance with the laboratory requirements—specifies which examinations will be covered. However, the selected modules should adhere to consensus standards where possible.

5.2.6.3. Reinforcement and Consolidation Tiers

For the purposes of this report, the Working Group assumed that the Reinforcement and Consolidation tiers are dedicated to forensic handwriting examinations.

This tiered approach to training allows for a process tailored to an individual based on criteria like the trainee's knowledge background, academic qualifications, and requirements for the individual or laboratory. The process gradually builds the range of knowledge, skills, and abilities required to undertake the specific role (be it handwriting expert or documents expert) and does so via competencies defined at three levels of achievement (see box 5.1).[566]

[566] J. C. Trinder, "Competency Standards—a Measure of the Quality of a Workforce," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVII, no. Part B6a (2008).

Box 5.1: Example of levels within the Reinforcement Tier in the tiered training process

Level 1—At this level, the trainees gain knowledge and understanding of the principles of forensic handwriting examination. They are introduced to the significance of handwritten features and characteristics, including use of specifically generated material (with ground truth known) to examine particular features encountered within handwriting. For example

- Types of handwriting, including natural, disguised, and traced/simulated;
- Neurophysiology of handwriting;
- Types of writing instruments;
- Levels and features of fluency; and
- Differences in individual character construction and combinations of characters.

Level 2—At this level, trainees apply their knowledge and understanding while they are introduced to the critical aspects of examining casework material, including

- Introduction to any relevant casework management systems employed by the organization;
- Understanding the purpose of submission and identifying what the potential outcomes of the examination may be;
- Determining that suitable and relevant material has been submitted and determining what other material may be required to complete the examination;
- Awareness of the other forensic opportunities that may be available, including other aspects of forensic document examination;
- Awareness of the impact of the examinations on other areas of forensic science, including potential contamination issues;
- Assessment of known and questioned material for internal consistency; and
- Awareness of potential sources of bias.

Level 3—At this level, the trainees demonstrate their depth of technical knowledge from exposure to the wide range of material submitted to the laboratory. This tier involves many separate examinations, potentially involving numerous case examples. The training includes

- Introduction to various types of material, including original and non-original documents;
- Introduction to case situations of varying size and complexity and how to manage them;
- Awareness of relevant databases, including the International Handwriting Information System, which includes international copybook styles and handwriting samples;
- Introduction to the relevant conclusion scale(s); and
- Preparation of forensic reports, including court comparison charts.

5.2.6.4. Reporting Tier

This is the final tier of the modular process. Reporting is the culmination of the training program and the decisive point in a trainee's progress. At the end of the training period, the trainee will undertake a series of competence assessments, including

- A review of the casework material examined during the training program to form a portfolio that can be assessed internally, and if appropriate, submitted for external scrutiny;
- Proficiency tests;
- Presentation skills tests, relating specifically to forensic handwriting comparisons;
- Report writing skills tests; and
- Moot court exercises.

5.2.6.5. Other Considerations

All aspects of training must be fully documented. As forensic science moves toward accreditation of the process and certification of the individual, this documentation will prove essential. The documentation should include the CV of all training officers; the syllabus of training; bibliography of reading material; internal test results; cases examined; instrumentation training; conferences, workshops, and outside classes attended; weekly reports from the training officer; and pre-training test results like color and form blindness.

The Working Group recognizes that some methods are not suitable for training and should not be considered acceptable. These methods include overreliance on distance learning, including periodic telephone conferencing and periodic meetings with training officers rather than regular face-to-face interactions. A training officer and trainee must have a routine and regular interface to accurately and fully assess development and progress.

Recommendation 5.3: The forensic document examiner community should develop a modular training program that consists of a publicly available standardized curriculum, as well as training and testing material.

To support this recommendation, the FDE community needs to explore funding options to establish a standardized modular-based competency assessment training program for forensic handwriting examination (see also Recommendation 5.1).

5.3. Final Competence Assessment and Certification

All FDE training before certification is currently undertaken in house, usually but not exclusively under the supervision of a training officer. Conventionally, a trainee is deemed competent by a series of final tests administered by the training officer. This process is not always open, transparent, or independent. Additionally, there are no standardized competency tests available for use by training officers, so each agency or private entity must develop its own tests or seek testing materials from others to use in its testing process. Once an FDE successfully completes training and passes all in-house competency tests, they may apply for certification by an external

certification body. An FDE's application for certification can be processed immediately following the successful completion of the training program, typically with the requirement that the individual is engaged in full-time forensic document practice. The Working Group recognizes that there is often a sizeable gap in time between the in-house testing process and the completion of the certification process, even if the application is submitted promptly upon eligibility. The Working Group suggests that the separate processes should be combined because (1) the in-house testing and certification processes have some redundant components, and (2) the in-house testing and certification goals are similar. Combining the competence testing and certification process into a single, externally accredited process may yield several benefits, including

- Assurance that FDEs passing the test are competent,
- Greater consistency in the level of assessment between candidates,
- Greater transparency in the independence of candidate testing,
- A consistent approach to the certification process,
- A higher number of candidates applying for certification, and
- Greater credibility for the certification process.

If pursued, this testing process should be rigorous and comprehensive and should be administered by an independent body comprising subject matter experts meeting current training standards, testing specialists, and other specialists as required. The comprehensive nature of the testing would require a significant amount of time. For example, testing for handwriting would necessarily include testing cursive, hand printing, numerals, disguise, numerous extrinsic factors, numerous intrinsic factors, simulation, tracing, writing transfer, foreign educated writers, and foreign language writing. Moot court would also be required because the ability to testify effectively in a competent and accurate manner is also a necessary skill for competent FDEs.

To ensure the appropriateness and independence of the testing process, an accredited certification organization should administer a single competency testing and certification process. This requires the formation of a new standard for testing the competency for FDEs. Any certification body that subsequently certifies the competency of an individual should do so based on this new standard.

To ensure a consistent approach to certification, all organizations that undertake the certification of individuals must be accredited to ISO/IEC 17024, "General requirements for bodies operating certification of persons," which is the only international set of accreditation requirements currently available.[567]

Recommendation 5.4: All forensic document examiners conducting handwriting examinations should be certified by a certifying body accredited to ISO/IEC 17024.

[567] See also the National Commission on Forensic Science (NCFS), *Views of the Commission: Facilitating Research on Laboratory Performance*.

5.4. Ongoing Education and Recertification

All certified FDEs must continue education or professional development per the requirements of their certifying organization. FDEs and others employed in the forensic sciences are subject to recertification. This recertification is also a standard for many other professional groups. Recertification allows FDEs to keep abreast of new technologies, legal requirements, and research in the field.

Several certifying boards in forensic science disciplines require those recertifying to document attendance at professional conferences and educational symposia, participation in educational workshops related to the field, and engagement in research activities, either through presentation of research papers at professional conferences and meetings within the discipline or publishing research results in peer-reviewed journals. The Working Group recognizes the importance of professional FDEs participating in educational workshops and conducting research within the discipline. However, mere attendance at professional conferences does not by itself provide for the FDE's continued education. Other disciplines require documented evidence in the form of continuing education credits (e.g., continuing education units, continuing medical education, continuing legal education). FDEs should provide documented evidence of attendance and participation at professional conferences, educational symposia, college coursework, and discipline-related workshops that have been pre-approved for credit as part of a structured recertification system. In addition, recertification and continuing education credit should be awarded for those FDEs who contribute to the professional literature through publications in peer-reviewed journals, presentations at professional conferences, and service on discipline-related boards and standards committees.

Furthermore, the Working Group recognizes the benefits of participating in routine proficiency testing (see section 4.2.6.2). This should form part of any continued professional development.

5.5. User Education—Communication of Expectations with the Legal Community

FDEs have voiced concern about the seemingly one-sided nature of procedural standardization, especially as it relates to conflicting comments, requests, and rulings by the legal profession. For example, FDEs have expressed frustration with the inconsistency of court rulings in which some judges have stated that they are only interested in definitive conclusions whereas other judges have stated that they would never accept or admit those experts claiming to be able to provide definitive conclusions. The lack of standardization in rulings is, of course, part of the judicial heritage. However, such widely expressed mutually exclusive positions create an untenable situation. The Working Group concluded an increase in direct communication between professional FDE groups and bar associations and between professional FDE groups and judicial gatherings would greatly help improve this disconnect.

The Working Group acknowledges that individual FDEs have previously provided presentations at various bar association meetings. Bar associations and the FDE community should encourage these contacts and increase their frequency. An open and continuous dialogue between attorneys and the FDE community should provide an atmosphere in which various concerns can be expressed, debated, and resolved.

Although judges and forensic scientists are part of the same process and strive for the ultimate goal of justice, they have limited opportunities where both communities can meet. To create

opportunities for communication and training, forensic scientists could reach out to organizations like the National Association of State Judicial Educators and attend other meetings where members of the judiciary and forensic scientists are present to discuss concerns and advancements. These interactions could provide a platform for in depth discussions on current issues affecting forensic science and forensic scientists.

Recommendation 5.5: Bar associations, judges' groups, and professional forensic document examiner organizations should collaborate to strengthen communication between the judiciary and forensic science communities for mutual benefit.

This page intentionally left blank.

6. Management

Introduction and Scope

An opinion proffered from a handwriting examination can directly impact a person's liberty, reputation, or financial health. With this in mind, previous chapters discussed how QA/QC (chapter 4) and education, training, and certification (chapter 5) help ensure the reliability of forensic handwriting examinations. This chapter delineates management's role in ensuring that these best practices are available to and followed by the FDE. In addition, this chapter discusses management's responsibility to provide FDEs with the appropriate tools, environment, and support to conduct their examinations.

This chapter applies to all FDEs regardless of laboratory size. To limit confusion, however, there are two concepts that warrant some explanation. First, management is used as a term for anyone more senior than an FDE in the organizational hierarchy who has some control of the work assignment. Second, when management is referred to in the context of a sole practitioner laboratory, this term also refers to the FDE. Naturally, the term management will not always be strictly synonymous with sole practitioners and may be more suited to a multi-person laboratory; however, sole practitioners should still consider how they can adjust their practice according to the topics discussed.

Other considerations for sole practitioner or small laboratories

Although the terms "management" and "quality manager" in this chapter refer to the FDE in a sole practitioner laboratory, these concepts do not always translate well to an environment where the manager and FDE are the same person. For example, section 6.3.1 deals with management's communication with the FDE, and section 6.7.3.1 considers management's leadership.

6.1. Management's Role in a Robust QA Program

Managing forensic handwriting examination service providers, from a single person laboratory to a large government agency, should involve consistent guiding principles. One key to appropriate management is the establishment and maintenance of a clearly defined QA program that is guided by international standards. Chapter 4 delineated how a robust QA program should be designed to ensure competency and ongoing proficiency, assist with laboratory accreditation, and regulate the review of policy and procedure manuals and examinations.

Accreditation and certification are also elements of a QA program that laboratories must consider. For a laboratory to prepare for accreditation, the most basic components include developing and implementing a procedure and quality manual, and participation in annual proficiency tests. Accreditation measures the quality system and how a laboratory meets those standards, whereas certification is a measure of an individual FDE's competency. Accreditation and certification should be used as a part of the quality program to increase the external review of the work conducted in the laboratory, and management must dedicate the appropriate resources (time, money, and support) to implement those activities.

The Working Group recognizes additional difficulties—financial and time costs—for smaller laboratories or sole practitioners to obtain accreditation. As this report went to press, the cost associated with gaining and maintaining accreditation was approximately $3,000 per year

(averaged over a 4-year accreditation cycle) for a sole practitioner laboratory.[568] Other costs, both in time and money, include developing and maintaining manuals, maintaining the quality program, and undertaking audits and technical reviews. As discussed in section 4.1, smaller laboratories or sole practitioners may benefit from working with accredited agencies to address some of the difficulties currently associated with accreditation for these service providers.

For those not yet accredited laboratories, management should seek to understand the advantages of accreditation. Management in smaller laboratories or sole practitioners should collaborate with larger laboratories and professional associations to become familiar with the accreditation process. The following is a sample list of actions for associations and larger laboratories to consider to assist laboratories who do not yet have accreditation:

- Provide workshops to discuss and encourage accreditation;
- Develop material explaining the purpose and benefits of being accredited that could be used to ensure continuity across the profession;
- Develop procedure and quality manual templates that could easily be adapted by a small or sole practitioner laboratory;
- Develop a template retainer agreement for civil FDEs that includes language about the use of a technical reviewer as a necessary part of the accreditation process; and
- Develop a network of FDEs who can provide technical reviews.

Recommendation 6.1: Management should dedicate appropriate resources to meet accreditation and certification requirements.

6.1.1. Additional Considerations for the Sole Practitioner

Sole practitioners are an important component in the justice system because they not only serve prosecutors but also provide services for criminal defense attorneys and attorneys seeking services for civil casework. The application of management and accreditation recommendations for sole practitioners, however, is particularly burdensome.

It is important to recognize that many recommendations will take time to implement and that it is unreasonable to demand that laboratories of all types satisfy these recommendations overnight. Equally, it is unreasonable to expect that laboratories will suspend work and cease providing services to the legal community until and unless these recommendations are implemented.

If further protection against errors is the goal, it should be the goal of all laboratories, large and small. Aiming to meet accreditation standards should therefore begin as soon as possible. It is anticipated that professional organizations will need to assist sole practitioners through the myriad requirements to meet international accreditation standards. Professional associations can

[568] Including fees for application, optional visit, full and interim assessments, accreditation maintenance and surveillance as well as participation in annual proficiency tests. Figure approximated based on discussions with various accreditation bodies: "ANSI National Accreditation Board (ANAB) home page," 2020, accessed May 11, 2020, https://anab.ansi.org/; American Association for Laboratory Accreditation (A2LA), "American Association for Laboratory Accreditation (A2LA) home page."; National Association of Testing Authorities (NATA), "National Association of Testing Authorities (NATA), Australia home page."

provide guidance documents and templates to their membership along with hosting workshops or other informational meetings for knowledge transfer.

6.2. Management's Role in Providing Appropriate Training

For FDEs to be reliable and accurate in their examinations, they must be trained by someone who has appropriate technical knowledge and an ability to mentor effectively. Management must provide the resources for training, including qualified and effective trainers. Although training methods should be tailored to the needs of the trainee(s), comprehensive training programs should adhere to consensus standards (see also section 5.2).

6.2.1. Continuing Education

Neglecting ongoing staff training and professional development can lead to failure to meet service goals and quality requirements because FDEs may not stay abreast of current laws, standards, techniques, technology, and procedures. Without continuing education, the reliability and accuracy of casework might be compromised (see also section 5.4).

Management has a responsibility to provide support for continued professional development that encompasses competency maintenance, skill enhancement, and other aspects of professional activities. Sources of training, internal or external to the laboratory, can include private industries and organizations, professional societies, mentors, training and academic institutions, and government agencies.

Management should maintain a continuing education record, including a description of the activity, format, date, and certificate or statement of completion.[569] Training and continuing professional development programs should undergo periodic external audits.

Management should also plan for any impact that continuing education and proficiency testing may have on case productivity. In addition to regular duties, practitioners will need time to pursue professional development and, if applicable, mentor trainees. Some agencies specify an annual training and continuing professional development budget for each FDE, which may include the provision of funds for travel and fees to complete outside learning opportunities. It is recommended that a forensic science laboratory establish a budget for training and continuing professional development.

Recommendation 6.2: Management must ensure appropriate resources are available and used for any initial, remedial, and ongoing competency training, including selection of qualified, effective trainers.

6.2.2. Assessment of Competency

Competency has typically been assessed through tests administered by the trainer at the completion of a trainee's training program. Although these tests provide key information on the trainee's competency, an independent assessment of competency has added benefits. After

[569] U.S. Department of Justice, *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*.

training, an FDE should pursue certification[570] administered by an independent and accredited board. The primary objective of a certification board is to administer comprehensive, validated tests that independently assess an applicant's competence. Certification must also be based on adherence to published best practices and standards in the discipline (e.g., SWGDOC, ASB). Certification boards also assess ongoing competence via recertification processes. It is critical that management support the independent confirmation of the new FDE's competency. (See section 5.3 for further discussion on competency assessment and certification.)

In the United States, questioned document certification involves demonstrating competency in handwriting and other aspects of questioned document examination, like ink comparisons, alterations to documents, printing processes, and indented impressions.[571] An FDE cannot currently be certified in questioned document examination if that individual only shows competence in handwriting examination.

6.3. Communication

6.3.1. Communication with FDEs

In multi-person laboratories, management should create an environment that encourages open communication between FDEs and their supervisors, the laboratory director, and the quality manager. This provides opportunities to identify and discuss problems FDEs may encounter and leads to greater transparency between management and FDEs. For example, open communication can help identify caseload and case management stress, interpersonal conflict, and business pressures. Management should ensure that FDEs have access to support services for emotional, work, or other related stresses or difficulties that could impact their well-being and work product.

Poor communication may consist of giving confusing or conflicting directions or demands, a failure to convey or obtain adequate information, lack of report writing skills, lack of teamwork, poor case documentation, departures from standard terminology, and conveying information in a way that could lead to bias in an examination. All these examples can adversely affect an FDE's performance. For instance, if management ambiguously conveys information about a task, the FDE could misinterpret the task. Furthermore, if management conveys information that is irrelevant and potentially biasing, this could lead to erroneous decision making. It is a delicate balance to limit communication to relevant information while still giving FDEs enough information to perform their tasks in an appropriate way.

6.3.2. Communication with Customers

The FDE must take steps to avoid unnecessary and potentially biasing case information. Management should, if possible, provide a case manager or an intermediary so that the proper examination can be made without task-irrelevant case information inadvertently influencing the examination process (see section 2.1). If an FDE is required to interact with the case submitter or client to ensure that the forensic examination is consistent with the request being made, it is

[570] Certification is not the same as accreditation. Certification assesses an individual's competence, whereas accreditation only assesses the laboratory as a system. U.S. Department of Justice, *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and Students*.

[571] ASTM International. *Standard Guide for Minimum Training Requirements for Forensic Document Examiners*.

important that only communication critical to the examination be provided to the FDE before analysis. Clients and case-submitters (e.g., attorneys) who interact directly with FDEs may require CIM training to understand the risk of bias when communicating task-irrelevant information.

6.3.3. Communication with Other Stakeholders

FDEs are likely to communicate with other FDEs, management, investigators, defense and prosecuting attorneys, administrative personnel, and other submitting parties. Although verbal communication is certainly important, communication via case documentation is imperative. Other FDEs can only adequately provide technical and administrative review with sufficient documentation and reporting. For instance, understanding the writing surface, writing instrument used, and other information can be critical for interpreting a questioned document. Additionally, understanding how the FDE compared the questioned document with a known sample can provide critical information in assessing if and how an error has occurred.

In criminal trials, FDEs should have the opportunity to discuss their findings with defense counsel and prosecutors. Discussing findings with both parties demonstrates transparency and impartiality. Management must also ensure that stakeholders are informed of deleterious events like mistakes, contaminated evidence, or anything else that could compromise the evidence or conclusions, even if they occur after testimony (see box 4.1).

6.4. Physical Environment

How a facility is designed and outfitted, including consideration for ergonomics and other human factors, can affect the FDE's ability to accomplish the needed tasks. Management must therefore consider how the work environment can create the best opportunity for an FDE to appropriately and successfully complete an examination and arrive at a proper conclusion.

The layout of a facility and the placement of instrumentation must be thought out carefully. Some individuals need a quiet place to work, whereas some can work in a noisy environment without problems. As such, the definition of a well-designed workplace is somewhat subjective and depends on the needs of the individual and the structure of the organization.

A laboratory's physical size will largely depend on the number of staff working in the space. Although space standards vary widely by organization, a range of 700 to 1,000 square feet per staff member offers a snapshot of the laboratory's potential size. A laboratory with fewer than 30 people may need about 1,000 square feet per staff member, whereas a larger facility of over 110 staff members may need only 720 square feet per staff member.[572]

Beyond space requirements and architectural design, management should also consider how the FDE's workspace can be maximized for safety, efficiency, and comfort.[573] Such considerations include the workstation and lighting.

---

[572] National Institute of Standards and Technology (NIST), *Forensic Science Laboratories: Handbook for Facility Planning, Design, Construction, and Relocation*, U.S. Department of Commerce (June 2013 2013), 14.

[573] National Institute of Standards and Technology (NIST), *Forensic Science Laboratories: Handbook for Facility Planning, Design, Construction, and Relocation*, 20.

6.4.1. Workstation

An ergonomically designed workstation may help to enhance the FDE's ability to organize and examine the handwriting and documents and can create a safer and more comfortable environment. For example, a large, slanted workstation can reduce neck strain caused by leaning over.[574]

6.4.2. Appropriate Lighting

Deciding on appropriate lighting requires a consideration of both the intensity and wavelength of the available light because both properties play a key role in the physics of how the eye can discriminate fine details and subtle color differentials. Natural daylight, typically from the north side of a building (in the northern hemisphere), tends to be considered the best[575] because the reflected or indirect light produces cool and controlled value shifts that help with color balance and consistency. Natural daylight helps the FDE assess subtle changes in ink and paper color. Daylight bulbs are readily available, which can provide a consistent and sufficiently intense light throughout the day. Furthermore, this lighting can reduce eyestrain.

6.5. Technical Environment

6.5.1. Equipment/Tools

A wide variety of examination tools are available to assist in the examination process, including basic magnification, microscopes, illumination devices, high-resolution scanning and photographic equipment, computer imaging software and hardware (i.e., fast processors to handle large image files and large, high-resolution monitors), spectral devices, and indentation detection devices.

Equipment that enhances the FDE's ability to see fine detail can be critical. For example, magnification allows the FDE to observe fine details of writing that might be missed with the naked eye, like regions where the pen has been lifted from the document and placed back down. Observing these features could play an important role in discerning the authenticity of a writing; therefore, management must provide the necessary equipment for a proper examination.

In addition to equipment required for the examination process, management should provide equipment to assist the FDE with research, report writing, and products intended to visually display the basis for any determinations the FDE makes.

6.5.2. Interfaces and Displays

Interfaces and displays can serve two distinct purposes in handwriting examination. First, they assist the FDE in assessing the evidence, for example by isolating images of comparable writing for creation of composite images. Second, visual representations of the examination process, such as images or illustrations, can assist the fact finder in understanding the basis for an

---

[574] Leaning over a desk (approximately 60°) can cause neck strain equivalent to a 60-pound weight hanging from the neck. (K. K. Hansraj, "Assessment of stresses in the cervical spine caused by posture and position of the head," *Surg Technol Int* 25, no. 25 (Nov 2014), https://www.ncbi.nlm.nih.gov/pubmed/25393825.)

[575] National Institute of Standards and Technology (NIST), *Forensic Science Laboratories: Handbook for Facility Planning, Design, Construction, and Relocation*, 14.

opinion. Images or illustrations must accurately reflect the evidence so that demonstrations are not misleading.

Recommendation 6.3: To provide the forensic document examiner with the best opportunity to make an appropriate examination, management must consider ergonomics of the work environment, including the influence of good lighting, sufficient workspace, and sufficient equipment.

6.6. Standardized Procedures

6.6.1. Manual Design

Laboratory manuals are a required part of accreditation as they provide the auditor with valuable information about laboratory processes and promote consistency in execution and application of particular methods. Regardless of accreditation requirements, all practitioners should have access to clearly designed manuals. Manuals relating equipment operation should describe the appropriate and effective use of that equipment and include logs that track maintenance performed on the equipment throughout its lifetime. Manuals should also provide a documented reference for how an FDE performs the various functions and uses equipment in the examination process. Management should support the development of appropriate and clearly designed manuals.

6.6.2. Procedure Design

Like manuals, formalized and documented procedures help ensure consistency in the way that FDEs approach their various tasks. For example, a well-designed checklist that is practical, precise, and designed for efficiency can streamline the examination process and reduce instances of neglected steps.[576]

Innovation and experimentation have been critical factors in developing new techniques and procedures in the field of forensic document examination from its inception. Task procedures must be designed and implemented in such a way that they do not stifle innovation.

6.7. Error Causation and Management

To identify, mitigate, and prevent human errors, management needs to understand their cause. Literature on human error describes many models of error causation. Such models include root cause analysis,[577] failure mode and effects analysis,[578] a management oversight risk tree,[579] the Human Factors Analysis and Classification System (HFACS),[580] and the "Swiss cheese" model.[581] If one considers the underlying assumptions regarding the nature and cause of error in

[576] A. Gawande, *The Checklist Manifesto* (London: Profile Books, 2010), 120.

[577] For example, in forensic science context, see Quattrone Center for the Fair Administration of Justice, *Guidelines for the Use of Root Cause Analysis (RCA) to Reduce Error and Improve Quality in Forensic Science Laboratories* (2014), https://www.nist.gov/system/files/documents/2016/11/22/guidelines_for_the_use_of_root_cause_analysis_to_reduce_error_and_improve_quality_in_forensic_science_labs.hollway.labmgmt.pdf.

[578] D. H. Stamatis, *Failure Mode and Effect Analysis: FMEA from Theory to Execution*, 2nd ed. (ASQ Quality Press, 2003).

[579] W. G. Johnson, "MORT: The Management Oversight and Risk Tree," *Journal of Safety Research* 7, no. 1 (1975).

[580] Shappell and Wiegmann, *The Human Factors Analysis and Classification System–HFACS.*

[581] J. Reason, "Human Error: Models and Management," *BMJ* 320, no. 7237 (Mar 18 2000), https://doi.org/10.1136/bmj.320.7237.768.

these models, there are at least six different perspectives to error investigation: (1) cognitive, (2) ergonomic, (3) behavioral, (4) medical, (5) psychosocial, and (6) organizational. Each perspective on human error investigation has its advantages, and many industries employ a multi-perspective approach.

The key assumption of these models is that human error in the workplace is not an isolated action of a given individual; rather, it is the result of a chain of events. This chain of events is described in James Reason's "Swiss cheese" model.[582] Reason's model assumes that all organizations have fundamental elements and systems that must work together harmoniously to achieve efficient and safe operations. Using this model of error causation, an error occurs when the "holes" from each "slice of cheese" are aligned.

Forensic analysis can be viewed as a complex system whose product is the interpretation of forensic evidence. Productive activities within a forensic unit require reliable, well-maintained equipment and a well-trained professional workforce. FDEs need good management and effective supervision, and managers need appropriate guidance, personnel, and funding to perform their duties. Accidents occur when there are breakdowns in the interaction among the components in the production process. These failures, depicted as holes in the metaphorical Swiss cheese slices, make the system more vulnerable to error.

This report considers four slices of Swiss cheese: (1) FDE actions, (2) FDE state, (3) management issues, and (4) organizational influences. FDE actions are the mistakes or violations by the examiner. Examiner state includes the physical and mental well-being of the examiner. Management issues relate to leadership, operational planning, problem correction, and management violations. Finally, organizational influences on the examiner relate to organizational structure, resource management, organizational climate, and operational processes.

Identifying weaknesses in a forensic system requires a two-stage approach: (1) a human error model to capture and organize the information and (2) an analysis of the examination process to identify the human and other factors that can affect the examination outcome. Using the Swiss cheese model, if an error has occurred, the investigation of the cause(s) starts with the FDE's actions, proceeds through the conditions that may have contributed to the error (including FDE state), and continues on to management actions and organizational oversights or failures.

6.7.1. Examiner Actions

At least two problematic FDE actions can lead to errors: mistakes and violations. Mistakes represent an FDE's actions performed with the intent to be correct but were in error. Violations, on the other hand, represent willful disregard of accepted practices. Management should take steps to identify when FDEs are performing actions that have the potential to result in mistakes and violations and appropriately address those actions. At the same time, management must foster a positive error culture by encouraging FDEs to acknowledge their own problematic actions and those others have committed, without the fear of retribution (see also section 6.8).

[582] Reason, "Human Error: Models and Management."

6.7.1.1. Decision-, Skill-, and Perception-Based Mistakes

Decisions are based primarily on three factors: information, knowledge, and experience. In handwriting examinations, information lies in the questioned and known writing samples, which must be of sufficient quality and quantity to compare and evaluate. In addition, the FDE should occasionally be provided with other information like the physical and mental state of the writer if the writing is distorted (e.g., a broken arm, medication, alcohol, or the lack of alcohol [for alcoholics]). These factors can all alter someone's natural writing.

In assessing evidence, the FDE applies training, background knowledge, and experience from comparing a broad range of questioned and known handwriting samples. When important information, knowledge, or experience is lacking, mistakes can occur. These errors typically present themselves as poorly executed procedures, improper choices, or the misinterpretation or misuse of relevant (or irrelevant) information.

Other mistakes occur with little or no conscious thought. For instance, frequent interruptions can disrupt the thought process. When resuming work after the disruption, an FDE may inadvertently skip steps in the examination. Such highly practiced and automatic behaviors are particularly affected by attention or memory failures. Distractions may lead to a loss of concentration, erroneous documentation, and other mistakes.

Additionally, mistakes can occur because of the way FDEs store and compare information. For instance, if notes are not taken contemporaneously to document the relevant features, FDEs must rely on their imperfect memory, which may distort their overall conclusions. These types of mistakes may present as failure to find target data, improper weight given to the data, failure to recognize disguise or distortion, and failure to compare enough corresponding features.

These types of mistakes may result in FDEs reaching conclusions not supported by the data or that are beyond their skill set, failing to search all exemplars, performing a hurried or insufficiently thorough examination, and improperly deeming a handwriting sample to be either suitable or unsuitable for comparison.

6.7.1.2. FDE Violations

A violation represents an action in which an FDE has intentionally or knowingly disregarded accepted practice. There are at least two types of violations: routine and exceptional. Commonly referred to as bending the rules, routine violations tend to be a habitual departure from procedures. This type of activity is often enabled by a system of supervision and management that tolerates minor departures from standard procedures. Just as some drivers may go 5 miles per hour over the speed limit and rarely suffer repercussions—and therefore believe it is not egregious—some FDEs may engage in shortcuts like not taking contemporaneous notes in the belief that they can accurately recall all their observations.

Akin to driving 30 miles per hour over the speed limit, an exceptional violation could occur when an FDE is pressured by a case submitter to reach a conclusion that is not supported by the evidence. Additional examples of exceptional violations include deeming a questioned document unsuitable for comparison to avoid having to compare it, disregarding aspects of the QA/QC process, intentionally misidentifying a questioned document, making an identification or exclusion of a handwriting sample that the FDE knows is not suitable for comparison, reporting

results without conducting a comparison, and coercing a verifier into agreeing with a rendered conclusion. Exceptional violations are particularly egregious; however, management must not condone any violation, regardless of its severity.

6.7.2. Examiner State

The second slice of the adapted Swiss cheese model relates to how the FDE's mental and physiological state, and physical or mental limitations, can affect performance. Examples include exhaustion, stress, anger, apprehension about reaching conclusions, boredom, complacency, distraction, expectancy, fatigue, overconfidence, peer pressure, and personal problems. If an FDE's condition interferes with duty performance, management should take appropriate action.

Situational factors like large backlogs could pressure FDEs to meet quotas or unrealistic turnaround times. Without appropriate management, FDEs could become more concerned with case output than their work quality. Shortcuts in the analysis and documentation of the handwriting evidence could lead an FDE to reach an inappropriate opinion. Management must take appropriate steps—such as being a buffer between the client and FDE and providing adequate staffing levels—so that large backlogs and other situational factors do not cause unnecessary stress and errors.

The FDE's physiological state can also affect the examination process. For example, the typical FDE usually bends over a desk or workbench and looks through a magnifier for long stretches of time. These working conditions can produce neck, back, and eye strain. Furthermore, glare from computer displays and the overwhelming number of comparisons can result in headaches or eyestrain.

Other factors bearing on an FDE's physiological state include illness, medication, alcohol and drug use, poor nutrition, injuries, lack of sleep, and poor sleep quality. For example, an FDE could be called to a crime scene in the middle of the night and then be expected to work a normal caseload the next day without rest. Management and FDEs should be aware of these risk factors and take steps to address and mitigate them.

Finally, physical and mental limitations should also be considered. Examples of limitations include deteriorating eyesight, inability to maintain competency, chronic psychological disorders, dyslexia, incompatible aptitude, and visual limitations such as poor acuity, poor contrast sensitivity, and color blindness. If an FDE cannot compensate for a physical or mental limitation, they may no longer be able to perform handwriting examinations. Management must take a role in identifying and mitigating such limitations. One way of identifying physical or psychological limitations is to implement a medical surveillance program that routinely checks for any health-related issues (e.g., declining eyesight) that might affect FDE performance.

6.7.3. Management Issues

The third layer of Swiss cheese in this adaptation of Reason's model[583] relates to management issues. The Working Group categorized these issues in relation to leadership, operational planning, problem correction, and management violations.

[583] Reason, "Human Error: Models and Management."

6.7.3.1. Leadership

Effective management includes effective leadership. An effective leader acts an advocate for the FDE; ensures appropriate training; sets a proper example; tracks and assesses job qualifications or skills; monitors work; provides appropriate feedback, mentoring, and incentives; maintains realistic expectations; and provides operational leadership.

The micromanagement of FDEs can delay decision making, restrict information flow, and diminish confidence and efficiency. Management should therefore provide sufficient oversight without becoming too controlling or more concerned with minute details than the accuracy of the work.

6.7.3.2. Operational Planning

Management is responsible for planning laboratory operations. Operational planning failures like not allowing adequate rest breaks; setting conflicting objectives, goals, or standards; giving unclear or conflicting assignments; and burdening FDEs with a heavy workload can all increase the risk of errors. Management should allocate casework to maintain productivity without causing frustration for FDEs. For example, the manager who assigns a large, complex case to a less experienced FDE may inadvertently set the FDE up for failure. Conversely, burdening the most efficient FDEs with excessive work can keep them from performing optimally and can limit the opportunities for less experienced FDEs to learn.

Scheduling should include breaks and should take caseloads and deadlines into account. FDEs with many rush (i.e., high priority) cases can feel overwhelmed, frustrated, and confused. Management should be aware of the risks in such cases and take precautions to prevent shortcuts or errors. Allowing FDEs to finish one batch of cases before assigning another batch can be helpful. Management should communicate risks to the courts to ensure that FDEs are given an appropriate and realistic amount of time to complete rush cases.

6.7.3.3. Problem Correction

If management is aware of problems, it should take action to correct these. Consistent failure to correct or discipline inappropriate behavior may foster a dysfunctional work environment. This caution also applies to issues associated with equipment and supplies. When necessary maintenance and repairs are overlooked or supplies do not meet specifications, errors can occur.

6.7.3.4. Management Violations

Management violations encompass the disregard of existing rules and regulations. An obvious example of poor management behavior is putting undue influence on an FDE to reach a desired result. A more subtle violation is permitting an unqualified or incompetent FDE to perform casework. Likewise, pushing FDEs to work unreasonably fast or encouraging them to bend the rules and not follow standard procedures in the interest of completing a case are also considered violations.

6.7.4. Organizational Influences

The fourth and final layer of Reason's Swiss cheese model[584] relates to organizational influences on FDE performance and error. Management must balance often competing goals of throughput, due diligence, and resources. These executive decisions are typically based on social, economic, and political input from outside the organization and on feedback from managers and workers within it. This report describes four areas of organizational influence: (1) organizational structure, (2) resource management, (3) organizational climate, and (4) operational processes.

6.7.4.1. Organizational Structure

Organizational structure refers to whether a laboratory is private (independent from law enforcement) or a branch of law enforcement. The NRC report[585] recommends that forensic agencies should be institutionally separated from law enforcement to ensure independence. The concern is that forensic scientists working within a law enforcement culture are at risk of aligning their own goals with those of investigators. The same concern can be raised with any FDEs who have direct contact with the client or case submitter if there are no processes in place to shield the FDE (or a reviewer) from irrelevant and potentially biasing information (see also section 2.1.1). Management should ensure that processes are in place to allow FDEs to assert their impartiality.

6.7.4.2. Resource Management

Resource management refers to the management, allocation, and maintenance of organizational resources, including human resource management (selection, training, staffing), budgets, logistics, and equipment design. Management decisions about such resources should focus on both quality and cost effectiveness. Unfortunately, quality improvements and training are often the first items to be cut during financial difficulty. Resource management issues include maintaining hiring, evaluation, and promotion policies; matching qualifications to job assignments; reducing costs and managing unfunded directives; providing logistical support; and making suitable equipment available.

6.7.4.3. Organizational Climate

Organizational climate refers to how members of an organization perceive and experience the culture of that organization. A negative organizational climate can adversely affect an FDE's performance. An FDE's experience of the organization can be influenced by components of the organizational structure, including the chain of command, delegation of authority and responsibility, communication channels, and formal accountability for actions. Agency policies that are ill-defined, adversarial, conflicting, or supplanted by unofficial rules and values can cause confusion, reduce quality, and lead to a negative organizational climate. Inaccessibility of upper management, inadequate accountability for actions, poorly defined or articulated organizational values, inappropriate allocation of resources, and unclear or conflicting assignments of responsibility can also lead to a negative organizational climate. Management is responsible for fostering a positive organizational climate.

---

[584] Reason, "Human Error: Models and Management."

[585] National Research Council (NRC), *Strengthening Forensic Science in the United States: A Path Forward*.

6.7.4.4. Operational Processes

Operational processes refer to decisions and processes that govern an organization's daily activities. Examples are SOPs and oversight methods that regulate the quality of work being completed. Management's role is to provide checks and balances to ensure staff follow standard procedures and do not take shortcuts. Management must monitor the risks through systems like checks to assess compliance with performance standards, objectives, and procedures; anonymous reporting systems; and a safety program with regular audits. Management must also avoid unduly enforcing productivity quotas beyond staff reach or compressing work completion schedules . These strategies may jeopardize the quality of the work completed.

6.8. Promoting Positive Error Culture

Errors are an inevitable part of human decision making. Rather than creating an environment of blame and hostility when these errors occur, management should see errors as an opportunity for learning, innovation, and resilience. In particular, by understanding how an error transpired, management and the FDE can improve processes to prevent the error from recurring. In this way, errors are managed to promote positive outcomes (i.e., promoting a positive error culture).

To create a positive error culture, management must foster a culture that promotes openness and acceptance—but not nonchalance—when errors are committed. To foster this culture, FDEs must feel safe and encouraged to report errors and have a sense that corrective actions will be taken when they do report errors.

Recommendation 6.4: Management should foster a culture in which it is understood that some human and system error is inevitable and that openness about errors will lead to improvements in practice.

6.9. Management's Role in CIM

The risk of contextual bias in forensic handwriting examination and methods for managing contextual information are discussed extensively in section 2.1. The Working Group calls on management to understand the risks associated with bias, to be informed on the latest research in the area, and to provide appropriate resources for the implementation of CIM procedures. Furthermore, the Working Group encourages management to facilitate FDE participation in research projects in this area.

6.10. Hiring FDEs

Little research has been conducted to test and validate what characteristics make a good forensic scientist. Typically, a candidate is hired based on an interview process and then begins training. There may be some value, however, in determining the types of people and skills best suited to perform handwriting examinations.[586] For example, employers could consider an applicant's spatial orientation abilities; ability to match incomplete patterns; cognitive, perceptual, and decision-making abilities; and comfort level with technology. Furthermore, it may be

[586] Julie Anne Schuck, *Personnel Selection in the Pattern Evidence Domain of Forensic Science* (2017).

advantageous for researchers to evaluate how a science or statistics degree and training in public speaking and technical writing may benefit the FDE.

7. Summary of Recommendations

Recommendation 2.1: The research community, in collaboration with forensic document examiners, should conduct research to study

- **The impact of various sources of contextual information on forensic handwriting examinations, and**
- **How to balance the risks of bias and information loss with respect to all levels of contextual information.**

Recommendation 2.2: Forensic document examiner laboratories performing handwriting examinations must use a contextual information management protocol, which must be documented within their quality management system.

Recommendation 2.3: Forensic document examiners must not report or testify, directly or by implication, that questioned handwriting has been written by an individual (to the exclusion of all others).

Recommendation 2.4: Forensic document examiners should collaborate with researchers to design and participate in "black box" and "white box" studies.

Recommendation 2.5: A forensic handwriting examination should be based on at least two mutually exclusive propositions relevant to the examination(s) requested. These propositions should be explicitly considered in the interpretation of the handwriting evidence and included in the conclusion, report, and testimony.

Recommendation 2.6: The forensic document examiner community should consider the claims made by forensic document examiners and then conduct empirical studies in collaboration with the research community to characterize the extent of scientific support for those claims.

Recommendation 2.7: The forensic document examiner community, in collaboration with researchers, should design and construct publicly available, large databases of representative handwriting features to facilitate research in and improve the accuracy of handwriting examination.

Recommendation 2.8: The forensic document examiner community should collaborate with the computer science and engineering communities to develop and validate applicable, user-friendly automated systems.

Recommendation 3.1: Whenever a handwriting examination is conducted, forensic document examiners should prepare reports as described in Recommendation 3.2, unless exempt by documented laboratory policy.

Recommendation 3.2: At a minimum, the forensic document examiner must include all the information listed below in the case record. Written reports must accurately and clearly detail all relevant aspects of analyses and comparisons. Unless this information is readily accessible by another mode (e.g., case record or report appendices), the written report should include

1. **Demographics: Submitter, forensic document examiner(s), laboratory, case identifier(s), or other information dictated by the laboratory.**
2. **Request for examination: What is being requested for each document.**
3. **Inventory of evidence: A listing or description of what documents are being submitted, their condition, and unambiguous identification of the items.**
4. **The curriculum vitae for each forensic document examiner.**
5. **A statement of case-related background information provided to the forensic document examiner(s).**
6. **A statement of propositions used in the evaluation of the evidence and a statement that if there are changes to the propositions, the opinion may change.**
7. **A statement of any assumptions made by the forensic document examiner and the basis for them and a statement that if there are changes in the assumptions, the opinion may change.**
8. **Methods: A listing of the instruments and methods used in the examination of the evidence, the range of possible conclusions, and a definition of terms.**
9. **Procedures: Specific step-by-step procedures for the examination of each document or set of documents and any deviations from established test methods.**
10. **Observations: A description of observed characteristics of each document or each set of documents and other bench notes.**
11. **Evaluations: The interpretation of the combined observations given each proposition.**
12. **Conclusions: A complete statement of the conclusions reached based on the observations and evaluations. When associations are made, the significance of the association should be communicated clearly and qualified properly. When exclusions are made, they shall be clearly communicated. When no conclusions are made, the reasons must be clearly stated.**
13. **Limitations: A statement of the limitations of the examination and the procedures.**

14. **Error rates: A statement of potential sources of error and, if available, relevant rates of error; if no relevant error rate is known by the laboratory, that fact should be disclosed.**

15. **Data: Charts, graphs, diagrams, or other data generated by the examination of the evidence as necessary for the proper understanding of the report.**

16. **Review of conclusions: If a review of conclusions occurred, whether a disagreement existed between the forensic document examiner and the reviewer.**

17. **Other statements required by the accreditation body or the laboratory.**

Recommendation 3.3: The forensic document examiner who conducts the examination and writes the report should be the one to testify in any proceeding.

Recommendation 3.4: Forensic document examiners must testify in a nonpartisan manner; answer questions from all counsel and the court directly, accurately, and fully; and provide appropriate information before, during, and after trial. All opinions must include an explanation of any data or information relied upon to form the opinion.

Recommendation 3.5: In testimony, a forensic document examiner must be prepared to describe the steps taken during the examination to reduce the risk of process, observational, and cognitive errors. The forensic document examiner must not state that errors are impossible.

Recommendation 3.6: Forensic document examiners must have a functional knowledge of the underlying scientific principles and research regarding handwriting examination, as well as reported error rates or other measures of performance, and be prepared to describe these in their testimony.

Recommendation 3.7: Demonstrative visual aids, when used, must be consistent with the report and anticipated verbal testimony. Aids must accurately represent the evidence, including both similarities and dissimilarities found in samples, and be prepared and presented in a manner that does not misrepresent, bias, or skew the information.

Recommendation 4.1a: Forensic document examiner laboratories* should be accredited to the current ISO/IEC 17025 standard by a recognized accrediting body.

***4.1b: In recognition of the practical constraints for sole practitioner laboratories to obtain accreditation, these laboratories should work toward meeting the requirements set forth in the current ISO/IEC 17025 standard and should become accredited when legitimate constraints are addressed.**

Recommendation 4.2: All forensic document examiner laboratories, whether accredited or not, must have a quality assurance and quality control system. This system should preferably align with the requirements of an international laboratory accreditation body.

Recommendation 4.3: The forensic document examiner community should collaborate with the research community and accreditation bodies to conduct and participate in studies to determine the optimal content and frequency of proficiency tests to properly evaluate forensic document examiners' ability to perform the range of tasks encountered in casework.

Recommendation 4.4: The forensic document examiner community should develop collaborative testing programs aimed at monitoring and providing performance improvement opportunities related to specific claims and sub-claims. The type, content, and frequency of these collaborative tests should be determined in consultation with the research community.

Recommendation 4.5: The forensic document examiner community should develop a framework for feedback-driven training, testing, and development based on ground-truth-known material.

Recommendation 4.6: Quality control procedures should include tracking of inconclusive and insufficient opinions. Test material should include these opinion categories.

Recommendation 5.1: To improve training, forensic document examiner professional organizations and practitioners should pursue both private and government funding like scholarships, grants, or loans to offset training costs.

Recommendation 5.2: Academia and professional forensic document examiner organizations should collaborate to develop trainer-skill workshops and classes.

Recommendation 5.3: The forensic document examiner community should develop a modular training program that consists of a publicly available standardized curriculum, as well as training and testing material.

Recommendation 5.4: All forensic document examiners conducting handwriting examinations should be certified by a certifying body accredited to ISO/IEC 17024.

Recommendation 5.5: Bar associations, judges' groups, and professional forensic document examiner organizations should collaborate to strengthen communication between the judiciary and forensic science communities for mutual benefit.

Recommendation 6.1: Management should dedicate appropriate resources to meet accreditation and certification requirements.

Recommendation 6.2: Management must ensure appropriate resources are available and used for any initial, remedial, and ongoing competency training, including selection of qualified, effective trainers.

Recommendation 6.3: To provide the forensic document examiner with the best opportunity to make an appropriate examination, management must consider ergonomics of the work environment, including the influence of good lighting, sufficient workspace, and sufficient equipment.

Recommendation 6.4: Management should foster a culture in which it is understood that some human and system error is inevitable and that openness about errors will lead to improvements in practice.

This page intentionally left blank.

References

Advisory Committee Notes. *Advisory Committee on Rules of Civil Procedure—May 1993.* (Washington, DC: 1993).

Federal Rules of Criminal Procedure Rule 16.

Airasian, Peter W., Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, and Merlin C. Wittrock. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* Edited by L. Anderson and D.R. Krathwohl. Boston: Allyn and Bacon, 2001.

Aitken, C. G. G., and D. A. Stoney. *The Use of Statistics in Forensic Science.* London: CRC Press, 1991.

Aitken, C., P. Roberts, and G. Jackson. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses.* Royal Statistical Society (2010). https://www.researchgate.net/publication/259088224_Fundamentals_of_Probability_and_Statistical_Evidence_in_Criminal_Proceedings_Guidance_for_Judges_Lawyers_Forensic_Scientists_and_Expert_Witnesses.

Al-Maadeed, Somaya. "Text-Dependent Writer Identification for Arabic Handwriting." *Journal of Electrical and Computer Engineering* 2012 (2012): 1-8. https://doi.org/10.1155/2012/794106.

Alaska Department of Public Safety Scientific Crime Detection Laboratory. "Alaska Crime Laboratory Quality Assurance Documents." n.d. https://dps.alaska.gov/Comm/CrimeLab/Quality-Assurance/QualityAssurance.

Alewijnse, L. C., C. E. van den Heuvel, and R. D. Stoel. "Analysis of Signature Complexity." *Journal of Forensic Document Examination* 21 (2011): 37–49.

Alston v. Shiver, 105 So. 785 (Flo. 1958).

American Association for Laboratory Accreditation (A2LA). "American Association for Laboratory Accreditation (A2LA) home page." 2020, accessed May 6, 2020, https://www.a2la.org.

American Bar Association (ABA). "Report of the ABA Criminal Justice Section's Ad Hoc Innocence Committee to Ensure the Integrity of the Criminal Process." In *Achieving Justice: Freeing the Innocent, Convicting the Guilty*, edited by P. C. Giannelli and M. Raeder. 47–62. Chicago: American Bar Association, 2006.

———. "Rule 1.3 Diligence—Comment." n.d., accessed May 7, 2020, https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_3_diligence/comment_on_rule_1_3.html.

American Board of Forensic Document Examiners (ABFDE). *Code of Ethics and Standard Practices.* (2014). https://www.abfde.org/htdocs/AboutABFDE/Ethics.pdf.

———. *Objectives for Training.* 2 ed., 2010.

American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB). *Potential Issues with Hair Comparison Testimony: Notification from the ASCLD/LAB Board of Directors to Interested Parties Concerning Potential Issues with Hair Comparison Testimony.* (2013).

American Society of Questioned Document Examiners (ASQDE). "Code of Ethics." n.d., accessed May 6, 2020, https://www.asqde.org/about/code_of_ethics.html.

ANSI-ASQ National Accreditation Board (ANAB). *General Requirements for the Competence of Testing and Calibration Laboratories.* ISO/IEC 17025:2017, AR 3125. ANAB, 2017.

ANSI National Accreditation Board (ANAB). "ANSI National Accreditation Board (ANAB) home page." 2020, accessed May 11, 2020, https://anab.ansi.org/.

———. "Forensic Accreditation." 2020, accessed May 6, 2020, https://anab.ansi.org/forensic-accreditation.

———. *Guiding Principles of Professional Responsibility for Forensic Service Providers and Forensic Personnel.* GD 3150. ANAB, 2018.

———. *ISO/IEC 17025:2005—Forensic Science Testing Laboratories Accreditation Requirements.* AR 3028. ANAB, 2017.

Arkansas State Crime Laboratory. "Quality Manuals." 2020, accessed May 6, 2020, https://www.crimelab.arkansas.gov/quality-manuals.

Association of Forensic Document Examiners (AFDE). "Code of Ethics." n.d., accessed May 6, 2020, https://afde.org/resources/AFDE_CODE-OF-ETHICS.pdf.

Association of Forensic Science Providers. "Standards for the Formulation of Evaluative Forensic Science Expert Opinion." *Science & Justice* 49, no. 3 (2009): 161–64.

ASTM International. *Standard Guide for Examination of Handwritten Items.* ASTM E2290-03. West Conshohocken, PA: ASTM International, 2003.

———. *Standard Guide for Minimum Training Requirements for Forensic Document Examiners.* ASTM E2388-11. West Conshohocken, PA: ASTM International, 2011.

———. *Standard Terminology Relating to the Examination of Questioned Documents.* ASTM E2195-02e1. West Conshohocken, PA: ASTM International, 2003.

Attorney General. *Memorandum for Heads of Department Components: Recommendation of the National Commission on Forensic Science; Announcement for NSFS Meeting Eleven.* (2016). https://www.justice.gov/opa/file/891366/download.

Austin Police Department, Quality Assurance Section. *Standard Operating Procedures.* (2016). https://web.archive.org/web/20170201010059/https://www.austintexas.gov/sites/default/files/files/Police/QA_Standard_Operating_Procedures_01-11-16.pdf.

Australia New Zealand Policing Advisory Agency (ANZPAA). "Education and Career Information." n.d., accessed May 6, 2020, https://www.anzpaa.org.au/forensic-science/forensic-sciences/education-and-career-information.

Balding, D. J. *Weight-of-Evidence for Forensic DNA Profiles.* Hoboken: John Wiley & Sons, 2005.

Bales, S. "Turning the Microscope Back on Forensic Scientists." *Litigation* 26, no. 2 (2000): 51-58. www.jstor.org/stable/29760125.

Ballantyne, K. N., G. Edmond, and B. Found. "Peer Review in Forensic Science." *Forensic Science International* 277 (Aug 2017): 66-76. https://doi.org/10.1016/j.forsciint.2017.05.020. https://www.ncbi.nlm.nih.gov/pubmed/28622536.

Baugh ex rel. Baugh v. Cuprum S.A.de C.V., 730 F.3d 701 (7th Cir. 2013).

Beacom, M. "A Study of Handwriting by Twins and Other Persons of Multiple Births." *Journal of Forensic Sciences* 5, no. 1 (1960): 121–31.

Behrendt, James E. "The Status of Training for Questioned Document Examiners in the United States." *Journal of Forensic Sciences* 34, no. 2 (1989): 366–70. https://doi.org/10.1520/jfs12645j.

Bird, C., B. Found, and D. Rogers. "Forensic Document Examiners' Skill in Distinguishing Between Natural and Disguised Handwriting Behaviors." *Journal of Forensic Sciences* 55, no. 5 (Sep 2010): 1291-5. https://doi.org/10.1111/j.1556-4029.2010.01456.x.

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl. *Handbook I: Cognitive Domain.* Taxonomy of Educational Objectives: The Classification of Educational Goals. London, WI: Longmans, Green and Co LTD, 1956.

Board of Forensic Document Examiners (BFDE). "Ethics." 2012, accessed May 6, 2020, https://www.bfde.org/ethics.html.

Boot, D. "An Investigation into the Degree of Similarity in the Handwriting of Identical and Fraternal Twins in New Zealand." *Journal of the American Society of Questioned Document Examiners* 1 (1998): 70–81.

Borsboom, D., G. J. Mellenbergh, and J. van Heerden. "The Concept of Validity." *Psychological Review* 111, no. 4 (Oct 2004): 1061-71. https://doi.org/10.1037/0033-295X.111.4.1061.

Borsboom, Denny, Gideon J. Mellenbergh, and Jaap van Heerden. "Functional Thought Experiments." *Synthese* 130, no. 3 (2002): 379-87. https://doi.org/10.1023/a:1014840616403.

Böttcher, C. .J. F. "Theory and Practice of Forensic Handwriting Comparison." *Tijdschift voor Strafrecht* 63 (1954): 77–131.

Bourjaily v. United States, 483 U.S. 171 (1987).

Bozza, Silvia, Franco Taroni, Raymond Marquis, and Matthieu Schmittbuhl. "Probabilistic Evaluation of Handwriting Evidence: Likelihood Ratio for Authorship." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57, no. 3 (2008): 329-41. https://doi.org/10.1111/j.1467-9876.2007.00616.x.

Brault, J. J., and R. Plamondon. "A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery." *IEEE Transactions on Systems, Man, and Cybernetics* 23, no. 2 (1993): 400-13. https://doi.org/10.1109/21.229453.

Bromwich, M.R. *Final Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room.* Fried, Frank, Harris, Shriver & Jacobson LLP (Washington, DC: 2007). http://www.hpdlabinvestigation.org/reports/070613report.pdf.

Browning, K. "Social Science Research on Forensic Science: The Story Behind One of NIJ's Newest Research Portfolios." *NIJ Journal* 275 (2015): 40–47. https://nij.gov/journals/275/Pages/social-science-forensics.aspx#.

Buckleton, J. S., C. M. Triggs, and C. Champod. "An Extended Likelihood Ratio Framework for Interpreting Evidence." *Science & Justice* 46, no. 2 (2006): 69-78. https://doi.org/10.1016/s1355-0306(06)71577-5.

Buckleton, J., C. M. Triggs, and S. J. Walsh. *Forensic DNA Evidence Interpretation.* Boca Raton, FL: CRC Press, 2005.

Budowle, B., M. C. Bottrell, S. G. Bunch, R. Fram, D. Harrison, S. Meagher, C. T. Oien*, et al.* "A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement." *Journal of Forensic Sciences* 54, no. 4 (Jul 2009): 798-809. https://doi.org/10.1111/j.1556-4029.2009.01081.x.

Bulacu, M. L. "Statistical Pattern Recognition for Automatic Writer Identification and Verification." Ph.D., University of Groningen, 2007.

Bulacu, M., L. Schomaker, and L. Vuurpijl. "Writer Identification Using Edge-Based Directional Features." *Proceedings of the Seventh International Conference on Document Analysis and Recognition* 2 (2003): 937-41. https://www.ai.rug.nl/~mbulacu/icdar2003-bulacu-schomaker-vuurpijl.pdf.

Butera, Karen D. "Seeing is Believing: A Practitioner's Guide to the Admissibility of Demonstrative Computer Evidence." *Cleveland State Law Review* 46, no. 3 (1998): 511-13.

Butler, J. *Advanced Topics in Forensic DNA Typing: Interpretation.* Oxford, UK: Academic Press, 2015.

Caligiuri, M. P., and L. A. Mohammed. "Chapter 3." In *The Neuroscience of Handwriting*. Boca Raton: CRC Press, 2012.

Carmines, E. G., and R. A. Zeller. *Reliability and Validity Assessment.* Sage Publications (London: 1979). https://academic.csuohio.edu/kneuendorf/c63111/C&ZSage.pdf.

Chaiken, S., A. Liberman, and A. H. Eagly. "Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context." In *Unintended Thoughts*, edited by J. S. Uleman and J. A. Bargh. 212–52. New York: The Guilford Press, 1989.

Champod, C., I. Evett, and B. Kuchler. "Earmarks as Evidence: A Critical Review." *Journal of Forensic Sciences* 46, no. 6 (2001). https://doi.org/10.1520/JFS15146J.

Christensen, A. M., C. M. Crowder, S. D. Ousley, and M. M. Houck. "Error and Its Meaning in Forensic Science." *Journal of Forensic Sciences* 59, no. 1 (Jan 2014): 123-6. https://doi.org/10.1111/1556-4029.12275.

Collaborative Testing Services (CTS), Inc. "Collaborative Testing Services (CTS), Inc. home page." 2020, accessed May 6, 2020, https://cts-forensics.com/.

———. *CTS Statement on the Use of Proficiency Testing Data for Error Rate Determinations.* Collaborative Testing Services (CTS), Inc. (Sterling, VA: March 30 2010). https://web.archive.org/web/20170117212801/http://www.ctsforensics.com/assets/news/CTSErrorRateStatement.pdf.

———. *Handwriting Examination Test No. 17-523/524 Summary Report.* (2018). https://cts-forensics.com/reports/3724_Web.pdf.

College of Physicians and Surgeons of British Columbia. *Laboratory Medicine Proficiency Testing Manual.* Diagnostic Accreditation Program (2018). https://www.cpsbc.ca/files/pdf/DAP-PT-Manual.pdf.

Conway, James V. P. "Evidential Documents." In *Scientific Examination of Questioned Documents*, edited by Ordway Hilton. CRC Press, 1959.

Cook, R. , I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert. "A Hierarchy of Propositions: Deciding Which Level to Address in Casework." *Science & Justice* 38, no. 4 (1998): 231–39. https://doi.org/10.1016/S1355-0306(98)72117-3.

Cooper, M. Q. "Practitioner's Guide: the Use of Demonstrative Exhibits at Trial." *Tulsa Law Journal* 34, no. 3 (1999): 567-78.

Curran, J. M., T. N. Hicks, and J. S. Buckleton. *Forensic Interpretation of Glass Evidence.* Boca Raton: CRC Press—Taylor & Francis Group, 2000.

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

Deng, L., G. Hinton, and B. Kingsbury. "New Types of Deep Neural Network Learning for Speech Recognition and Related Applications: An Overview." 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, 2013.

District of Columbia Department of Forensic Sciences. "Open Government and FOIA - DFS." n.d., accessed May 6, 2020, https://dfs.dc.gov/page/open-government-and-foia-dfs.

Dror, I. E., C. Champod, G. Langenburg, D. Charlton, H. Hunt, and R. Rosenthal. "Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison." *Forensic Science International* 208, no. 1-3 (May 20 2011): 10-7. https://doi.org/10.1016/j.forsciint.2010.10.013. https://www.ncbi.nlm.nih.gov/pubmed/21129867.

Dror, I. E., D. Charlton, and A. E. Peron. "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications." *Forensic Science International* 156, no. 1 (Jan 6 2006): 74-8. https://doi.org/10.1016/j.forsciint.2005.10.017. https://www.ncbi.nlm.nih.gov/pubmed/16325362.

Dror, I. E., W. C. Thompson, C. A. Meissner, I. Kornfield, D. Krane, M. Saks, and M. Risinger. "Letter to the Editor—Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making." *Journal of Forensic Sciences* 60, no. 4 (Jul 2015): 1111-2. https://doi.org/10.1111/1556-4029.12805. https://www.ncbi.nlm.nih.gov/pubmed/26088016.

Dror, I. E., K. Wertheim, P. Fraser-Mackenzie, and J. Walajtys. "The Impact of Human-Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts." *Journal of Forensic Sciences* 57, no. 2 (Mar 2012): 343-52. https://doi.org/10.1111/j.1556-4029.2011.02013.x.

Dror, Itiel E. "Cognitive Neuroscience in Forensic Science: Understanding and Utilizing the Human Element." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370, no. 1674 (Aug 5 2015). https://doi.org/10.1098/rstb.2014.0255. https://www.ncbi.nlm.nih.gov/pubmed/26101281.

———. "A Hierarchy of Expert Performance." *Journal of Applied Research in Memory and Cognition* 5, no. 2 (2016): 121-27. https://doi.org/10.1016/j.jarmac.2016.03.001.

———. "Human Expert Performance in Forensic Decision Making: Seven Different Sources of Bias." *Australian Journal of Forensic Sciences* 49, no. 5 (2017): 541-47. https://doi.org/10.1080/00450618.2017.1281348.

———. "The Paradox of Human Expertise: Why Experts Can Get It Wrong." Chap. 177-188 In *The Paradoxical Brain*, edited by N. Kapur. Cambridge: Cambridge University Press, 2011.

———. "Practical Solutions to Cognitive and Human Factor Challenges in Forensic Science." *Forensic Science Policy & Management: An International Journal* 4, no. 3-4 (2014): 105-13. https://doi.org/10.1080/19409044.2014.901437.

Dror, Itiel E., and David Charlton. "Why Experts Make Errors." *Journal of Forensic Identification* 56, no. 4 (2006): 600–16.

Dror, Itiel E., and S. A. Cole. "The Vision in 'Blind' Justice: Expert Perception, Judgment, and Visual Cognition in Forensic Pattern Recognition." *Psychonomic Bulletin & Review* 17, no. 2 (Apr 2010): 161-7. https://doi.org/10.3758/PBR.17.2.161. https://www.ncbi.nlm.nih.gov/pubmed/20382914.

Dror, Itiel E., and G. Hampikian. "Subjectivity and Bias in Forensic DNA Mixture Interpretation." *Science & Justice: Journal of the Forensic Science Society* 51, no. 4 (Dec 2011): 204-8. https://doi.org/10.1016/j.scijus.2011.08.004.

Dror, Itiel E., B. M. McCormack, and J. Epstein. "Cognitive Bias and its Impact on Expert Witnesses and the Court." *The Judges' Journal* 54, no. 4 (2015): 8-14. https://www.americanbar.org/groups/judicial/publications/judges_journal/2015/fall/cognitive_bias_and_its_impact_on_expert_witnesses_and_the_court/.

Dror, Itiel E., and J. L. Mnookin. "The Use of Technology in Human Expert Domains: Challenges and Risks Arising from the Use of Automated Fingerprint Identification Systems in Forensic Science." *Law, Probability and Risk* 9, no. 1 (2010): 47-67. https://doi.org/10.1093/lpr/mgp031.

Dror, Itiel E., and Daniel C. Murrie. "A Hierarchy of Expert Performance Applied to Forensic Psychological Assessments." *Psychology, Public Policy, and Law* 24, no. 1 (2018): 11-23. https://doi.org/10.1037/law0000140.

Dror, Itiel E., Ailsa E. Péron, Sara-Lynn Hind, and David Charlton. "When Emotions Get the Better of Us: the Effect of Contextual Top-Down Processing on Matching Fingerprints." *Applied Cognitive Psychology* 19, no. 6 (2005): 799-809. https://doi.org/10.1002/acp.1130.

Dror, Itiel E., and R. Rosenthal. "Meta-Analytically Quantifying the Reliability and Biasability of Forensic Experts." *Journal of Forensic Sciences* 53, no. 4 (Jul 2008): 900-3. https://doi.org/10.1111/j.1556-4029.2008.00762.x.

Durina, M., and M. Caligiuri. "The Determination of Authorship from a Homogenous Group of Writers." *Journal of the American Society of Questioned Document Examiners* 12, no. 2 (2009): 77–90.

Durose, M. R., A. M. Burch, K. Walsh, and E. Tiry. *Publicly Funded Forensic Crime Laboratories: Quality Assurance Practices, 2014.* NCJ 250152. U.S. Department of Justice, 2016.

Edmond, G., K. Martire, R. Kemp, D. Hamer, B. Hibbert, A. Ligertwood, G. Porter*, et al.* "How to Cross-Examine Forensic Scientists: A Guide for Lawyers." *Australian Bar Review* 39 (2014): 174–75.

Edwards, K., and E. E. Smith. "A Disconfirmation Bias in the Evaluation of Arguments." *Journal of Personality and Social Psychology* 71, no. 1 (1996): 5–24.

Eiserman, H. W., and M. R. Hecker. "FISH-computers in handwriting examinations." 44th Annual Meeting of the American Society of Questioned Document Examiners, Savannah, GA, 1986.

Ellen, D. *Scientific Examination of Documents: Methods and Techniques.* 3rd ed. Boca Raton: CRC Press—Taylor & Francis Group, 2006.

Ericsson, K. A. "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance." Chap. 38 In *The Cambridge Handbook of Expertise and Expert Performance*, edited by N. Charness K. A. Ericsson, P. J. Feltovich, and R. R. Hoffman. 685–706: Cambridge University Press, 2006.

———. *Peak: Secrets from the New Science of Expertise.* Boston: Houghton Mifflin Harcourt, 2016.

Ericsson, K. A., R. T. Krampe, and C. Tesch-Romer. "The Role of Deliberate Practice in the Acquisition of Expert Performance." *Psychological Review* 100, no. 3 (1993): 363–406.

European Network of Forensic Science Institutes (ENFSI). *Best Practice Manual for the Forensic Examination of Handwriting.* (November 2015). https://enfsi.eu/wp-content/uploads/2016/09/2._forensic_examination_of_handwriting_0.pdf.

———. *Best Practice Manual for the Forensic Examination of Handwriting.* ENFSI-BPM-FHX-01. June, 2018.

———. *Guideline for Evaluative Reporting in Forensic Science.* (2015). https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.

European Network of Forensic Science Institutes (ENFSI) Standing Committee for Quality and Competence. *Performance Based Standards for Forensic Science Practitioners.* I3 (d). 2004.

Evett, I. E. "Interpretation: a Personal Odyssey." In *The Use Of Statistics In Forensic Science*, edited by C. G. G. Aitken and D. A. Stoney. London: CRC Press, 1991.

Evett, I. W., G. Jackson, and J. A. Lambert. "More on the Hierarchy of Propositions: Exploring the Distinction Between Explanations and Propositions." *Science & Justice* 40, no. 1 (2000): 3-10. https://doi.org/10.1016/s1355-0306(00)71926-5.

Evett, I. W., and J. A. Lambert. "The Interpretation of Refractive Index Measurements. III." *Forensic Science International* 20, no. 3 (1982): 237-45. https://doi.org/10.1016/0379-0738(82)90123-2.

Evett, I. W., and Weir. B. S. *Interpreting DNA Evidence.* Sunderland, MA: Sinauer, 1998.

Evett, I. W. V. "The Interpretation of Refractive Index Measurements." *Forensic Science* 9 (1977): 209-17. https://doi.org/10.1016/0300-9432(77)90093-0.

F.R.E. 703.

F.R.E. 702.

Federal Bureau of Investigation. "FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review." News release, 2015, https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review.

Federal Rule of Criminal Procedure 16(b)(1)(C).

Federal Rule of Criminal Procedure 16(b)(1)(B).

Federal Rules of Civil Procedure Rule 26(a)(2)(B).

Federal Rules of Civil Procedure 26(b)(4)(D).

Federal Rules of Criminal Procedure Rule 16(a)(1)(F).

Federal Rules of Evidence 104(a).

Festinger, L. *A Theory of Cognitive Dissonance.* Stanford, CA: Stanford University Press, 1957.

Forensic Specialties Accreditation Board (FSAB). "Forensic Specialties Accreditation Board (FSAB) home page." Updated April 5, 2020, 2019, http://thefsab.org/.

Found, B., and J. Ganas. "The Management of Domain Irrelevant Context Information in Forensic Handwriting Examination Casework." *Science & Justice: Journal of the Forensic Science Society* 53, no. 2 (Jun 2013): 154-8. https://doi.org/10.1016/j.scijus.2012.10.004. https://www.ncbi.nlm.nih.gov/pubmed/23601722.

Found, B., and D. Rogers. "Contemporary Issues in Forensic Handwriting Examination: A Discussion of Key Issues in the Wake of the Starzecpyzel Decision." *Journal of Forensic Document Examination* 8 (1995): 1–31.

———. "The Initial Profiling Trial of a Program to Characterize Forensic Handwriting Examiners' Skill." *Journal of the American Society of Questioned Document Examiners* 6, no. 2 (2003): 72–81.

———. "The Probative Character of Forensic Handwriting Examiners' Identification and Elimination Opinions on Questioned Signatures." *Forensic Science International* 178, no. 1 (Jun 10 2008): 54-60. https://doi.org/10.1016/j.forsciint.2008.02.001.

———. "Problem Types of Questioned Handwritten Text for Forensic Document Examiners." International Graphonomics Society Biennial Conference, Salerno, Italy, Editrice Zona, Arezzo, Italy, June 26–29 2005.

Found, B., D. Rogers, and A. Herkt. "The Skill of a Group of Document Examiners in Expressing Handwriting and Signature Authorship and Production Process Opinions." *Journal of Forensic Document Examination* 14 (2001): 15–30.

Found, B., D. Rogers, V. Rowe, and D. Dick. "Statistical Modelling of Experts' Perceptions of the Ease of Signature Simulation." *Journal of Forensic Document Examination* 11 (1998): 73–99.

Found, B., J. Sita, and D. Rogers. "The Development of a Program for Characterising Forensic Handwriting Examiners' Expertise: Signature Examination Pilot Study." *Journal of Forensic Document Examination* 12 (1999): 69–80.

Found, Bryan J., and Carolyne Bird. "The Modular Forensic Handwriting Method—2016 Version." *Journal of Forensic Document Examination* 26 (2016): 7-83. https://doi.org/10.31974/jfde26-7-83.

Franke, K., L. Schomaker, L. Vuurpijl, and S. Giesler. "FISH-New: A Common Ground for Computer-Based Forensic Writer Identification." *Forensic Science International* 136, no. S1-S432 (2003): 84.

Franke, K.Y., L.R.B. Schomaker, C. Veenhuis, L. G. Vuurpijl, M. Erp, and I. van Guyon. "WANDA: A Common Ground for Forensic Handwriting Examination and Writer Identification." *ENFHEX News—Bulletin of the European Network of Forensic Handwriting Experts* 1, no. 4 (2001): 23-47. https://www.academia.edu/26020856/WANDA_A_common_ground_for_forensic_handwriting_examination_and_writer_identification.

Fraser-Mackenzie, P. A., I. E. Dror, and K. Wertheim. "Cognitive and Contextual Influences in Determination of Latent Fingerprint Suitability for Identification Judgments." *Science & Justice: Journal of the Forensic Science Society* 53, no. 2 (Jun 2013): 144-53. https://doi.org/10.1016/j.scijus.2012.12.002.

Frey, D., S. Schultz-Hardt, I. von Haeften, and H. Bresnitz. Information seeking under suboptimal conditions: The importance of time pressure and complexity for selective exposure to information. 2000. University of Munich.

Frey, D., and S. Schulz-Hardt. "Confirmation Bias in Group Information Seeking and Its Implications for Decision Making in Administration, Business and Politics." Chap. 4 In *Social Influence in Social Reality: Promoting Individual and Social Change*, edited by F. Butera and G. Mugny. 53–74, 2001.

Frey, D., D. Stahlberg, and A. Fries. "Information Seeking of High- and Low-Anxiety Subjects After Receiving Positive and Negative Self-Relevant Feedback." *Journal of Personality* 54, no. 4 (Dec 1986): 694-703. https://doi.org/10.1111/j.1467-6494.1986.tb00420.x.

Frey, Dieter. "The Effect of Negative Feedback About Oneself and Cost of Information on Preferences for Information About the Source of this Feedback." *Journal of Experimental Social Psychology* 17, no. 1 (1981): 42-50. https://doi.org/10.1016/0022-1031(81)90005-6.

———. "Postdecisional Preference for Decision-Relevant Information as a Function of the Competence of its Source and the Degree of Familiarity with this Information." *Journal of Experimental Social Psychology* 17, no. 1 (1981): 51-67. https://doi.org/10.1016/0022-1031(81)90006-8.

———. "Recent Research on Selective Exposure to Information." *Advances in Experimental Social Psychology* 19 (1986): 41-80. https://doi.org/10.1016/S0065-2601(08)60212-9.

Frey, Dieter, and Marita Rosch. "Information Seeking after Decisions: The Roles of Novelty of Information and Decision Reversibility." *Personality and Social Psychology Bulletin* 10, no. 1 (1984): 91-98. https://doi.org/10.1177/0146167284101010.

Frey, Dieter, and Dagmar Stahlberg. "Selection of Information after Receiving more or Less Reliable Self-Threatening Information." *Personality and Social Psychology Bulletin* 12, no. 4 (1986): 434-41. https://doi.org/10.1177/0146167286124006.

Frey, Dieter, and Robert A. Wicklund. "A Clarification of Selective Exposure." *Journal of Experimental Social Psychology* 14, no. 1 (1978): 132-39. https://doi.org/10.1016/0022-1031(78)90066-5.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Gamble, D. J. "The Handwriting of Identical Twins." *Canadian Society of Forensic Science Journal* 13, no. 1 (1980): 11-30. https://doi.org/10.1080/00085030.1980.10757337.

Gantz, D. T., and M. A. Walch. "FLASH ID Handwriting Derived Biometric Analysis Software." NIST Measurement Science and Standards in Forensic Handwriting Analysis Conference, 2013.

Gawande, A. *The Checklist Manifesto.* London: Profile Books, 2010.

General Electric Co. v. Joiner, 522 U.S. 136 (1997).

Good, I. .J. *Good Thinking: Foundations of Probability and Its Applications.* Minneapolis, MN: University of Minnesota Press, 1983.

———. *Probability and the Weighing of Evidence.* London: Charles Griffin and Company, 1950.

———. "Weight of Evidence and the Bayesian Likelihood Ratio." In *The Use of Statistics in Forensic Science*, edited by C. G. G. Aitken and D. A. Stoney. London: CRC Press, 1991.

Graham, Michael H. "Real and Demonstrative Evidence, Experiments and Views." *Criminal Law Bulletin* 46, no. 4 (2010): 792. https://ssrn.com/abstract=1885714.

Hájek, A., "Interpretations of Probability," E. N. Zalta ed. *The Stanford Encyclopedia of Philosophy*, 2012, https://plato.stanford.edu/archives/win2012/entries/probability-interpret.

Hall, L. J., and E. Player. "Will the Introduction of an Emotional Context Affect Fingerprint Analysis and Decision-Making?". *Forensic Science International* 181, no. 1-3 (Oct 25 2008): 36-9. https://doi.org/10.1016/j.forsciint.2008.08.008.

Hamilton, Alexander, James Madison, and John Jay. *The Federalist: A Collection of Essays Written in Favour of the New Constitution as Agreed Upon by the Federal Convention.* (1788). https://www.loc.gov/item/09021562/.

Hansraj, K. K. "Assessment of stresses in the cervical spine caused by posture and position of the head." *Surg Technol Int* 25, no. 25 (Nov 2014): 277-9. https://www.ncbi.nlm.nih.gov/pubmed/25393825.

Hardavella, G., A. Aamli-Gaagnat, N. Saad, I. Rousalova, and K. B. Sreter. "How to Give and Receive Feedback Effectively." *Breathe (Sheff)* 13, no. 4 (Dec 2017): 327-33. https://doi.org/10.1183/20734735.009917.

Harris County Institute of Forensic Sciences. *Non Conformity, Corrective and Preventive Action Procedure.* Procedure No. QP08.0007. 2016.

Harrison, D., T. M. Burkes, and D. P. Seiger. "Handwriting Examination: Meeting the Challenges of Science and the Law." *FBI Forensic Science Communications* 11, no. 4 (2009). https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2009/review/2009_10_review02.htm.

Harrison, W. R. *Suspect Documents: Their Scientific Examination.* London: Sweet & Maxwell Limited, 1958.

Hattie, John, and Helen Timperley. "The Power of Feedback." *Review of Educational Research* 77, no. 1 (2016): 81-112. https://doi.org/10.3102/003465430298487.

Hicks, T., A. Biedermann, J. A. de Koeijer, F. Taroni, C. Champod, and I. W. Evett. "The Importance of Distinguishing Information From Evidence/Observations When Formulating Propositions." *Science & Justice: Journal of the Forensic Science Society* 55, no. 6 (Dec 2015): 520-5. https://doi.org/10.1016/j.scijus.2015.06.008. https://www.ncbi.nlm.nih.gov/pubmed/26654089.

Hilton, O. *Scientific Examination of Questioned Documents.* Revised ed. New York: Elsevier North Holland, Inc., 1982.

HN Proficiency Testing, Inc,. "What is Proficiency Testing?", Updated October 24, 2019, 2001, https://www.hn-proficiency.com/profi.htm.

Hoffrage, U., and G. Gigerenzer. "Using Natural Frequencies to Improve Diagnostic Inferences." *Academic Medicine* 73, no. 5 (1998): 538–40. https://doi.org/10.1097/00001888-199805000-00024.

Holmes, Lindsay. "Handwriting Instruction in Canadian Schools as Prescibed by Provincial and Territorial Ministries of Education." *Canadian Society of Forensic Science Journal* 43, no. 1 (2010): 9-15. https://doi.org/10.1080/00085030.2010.10757616.

Houck, M., and J. A. Siegel. *Fundamentals of Forensic Science.* Burlington, MA: Academic Press, 2009.

Howard, M., and J. Barnum. "Bringing Demonstrative Evidence in from the Cold: The Academy's Role in Developing Model Rules." *Temple Law Review* 88, no. 3 (2016): 513-49. https://ssrn.com/abstract=2710182.

Huber, R. A. "Expert Witnesses." *Criminal Law Quarterly* 2, no. 3 (1959): 276–95.

Huber, R. A., and A. M. Headrick. *Handwriting Identification: Facts and Fundamentals.* Boca Raton: CRC Press LLC, 1999.

Hunt, T. R. "Scientific Validity and Error Rates: A Short Response to the PCAST Report." *Fordham Law Review Online* 86, no. 14 (2017). https://ir.lawnet.fordham.edu/flro/vol86/iss1/14.

Idaho State Police. "Welcome to Idaho State Police Forensic Services." Updated April 13, 2020, https://isp.idaho.gov/forensics/.

Indiana State Police Laboratory. *Quality Assurance Manual.* (2016). https://www.in.gov/isp/labs/files/Lab_QA_Manual_03-16-16.pdf.

International Laboratory Accreditation Cooperation (ILAC). "About ILAC." 2019, accessed May 6, 2020, https://ilac.org/about-ilac/.

———. *Modules in a Forensic Science Process.* ILAC-G19:08/2014. 2014.

International Organization for Standardization (ISO). *Conformity Assessment—General Requirements for Proficiency Testing.* ISO/IEC 17043:2010. Geneva, Switzerland: ISO, 2010.

———. *General Requirements for the Competence of Testing and Calibration Laboratories.* ISO/IEC 17025:2005. Geneva, Switzerland: ISO, 2005.

———. "ISO/IEC 17025:2017(en) General Requirements for the Competence of Testing and Calibration Laboratories." 2017, https://www.iso.org/obp/ui/#iso:std:iso-iec:17025:ed-3:v1:en.

———, "New Edition of ISO/IEC 17025 Just Published," *ISO News*, December 1, 2017, https://www.iso.org/news/ref2250.html.

Jackson, G., C. Aitken, and Roberts P. *Case Assessment and Interpretation of Expert Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses.* Royal Statistical Society (London: 2014). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.723.2831&rep=rep1&type=pdf.

Jackson, R., and R. Roesch, eds. *Learning Forensic Assessment: Research and Practice*. 2nd ed. New York: Routledge, 2016.

Johnson, M. E., T. W. Vastrick, M. Boulanger, and E. Schuetzner. "Measuring the Frequency Occurrence of Handwriting and Handprinting Characteristics." *Journal of Forensic Sciences* 62, no. 1 (January 2017): 142-63. https://doi.org/10.1111/1556-4029.13248. https://www.ncbi.nlm.nih.gov/pubmed/27864959.

Johnson, W. G. "MORT: The Management Oversight and Risk Tree." *Journal of Safety Research* 7, no. 1 (1975): 4-15.

Joint Committee for Guides in Metrology (JCGM). "Online abridged version of the International vocabulary of metrology—Basic and general concepts and associated terms (VIM)." Updated April 29, 2017, https://jcgm.bipm.org/vim/en/

Jonas, E., S. Schulz-Hardt, D. Frey, and N. Thelen. "Confirmation Bias in Sequential Information Search after Preliminary Decisions: an Expansion of Dissonance Theoretical Research on Selective Exposure to Information." *Journal of Personality and Social Psychology* 80, no. 4 (Apr 2001): 557-71. https://doi.org/10.1037//0022-3514.80.4.557.

Jones, A. W. "The Distribution of Forensic Journals, Reflections on Authorship Practices, Peer-Review and Role of the Impact Factor." *Forensic Science International* 165, no. 2-3 (Jan 17 2007): 115-28. https://doi.org/10.1016/j.forsciint.2006.05.013.

Jones, J. P. "The Future State of Handwriting Examinations: A Roadmap to Integrate the Latest Measurement Science and Statistics." AAFS Annual Meeting, Seattle, WA, February 20, 2014 2014.

Kadane, J. B. "Appropriate Statistics." In *Forensic Science Research Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature*, edited by E. G. Bartrick and M. A. Floyd. National Institute of Justice, 2015.

Kahneman, Daniel, and Amos Tversky. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3, no. 3 (1972): 430-54. https://doi.org/10.1016/0010-0285(72)90016-3.

Kam, Moshe, Kishore Gummadidala, Gabriel Fielding, and Robert Conn. "Signature Authentication by Forensic Document Examiners." *Journal of Forensic Sciences* 46, no. 4 (2001). https://doi.org/10.1520/jfs15062j.

Kang, T. Y., and J. Lee. "Multiform Korean handwriting authentication by forensic document examiners." *Forensic Science International* 255 (Oct 2015): 133-6. https://doi.org/10.1016/j.forsciint.2015.07.002.

Karatzoglou, Alexandros, and Balázs Hidasi. "Deep Learning for Recommender Systems." Proceedings of the Eleventh ACM Conference on Recommender Systems, Como, Italy, Association for Computing Machinery, August 2017 2017.

Karson, Michael, and Lavita Nadkarni. *Principles of Forensic Report Writing.* 2013. doi:10.1037/14182-000.

Kaye, D. H. "Identification, Individualization and Uniqueness: What's the Difference?". *Law, Probability and Risk* 8, no. 2 (2009): 85-94. https://doi.org/10.1093/lpr/mgp018.

———. "Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies." *Brooklyn Law Review* 75 (2010): 1163-86. https://elibrary.law.psu.edu/cgi/viewcontent.cgi?article=1015&context=fac_works.

Kaye, D. H., D. E. Bernstein, J. L. Mnookin, R. D. Friedman, and J. H. Wigmore. *The New Wigmore: A treatise on Evidence: Expert Evidence.* Austin, TX: Aspen Publishers, 2011.

Kaye, D.H., and I. M. Ellman. "Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?". *New York University Law Review* 54 (1979): 1131-62.

Kaye, David. "The Laws of Probability and the Law of the Land." *The University of Chicago Law Review* 47, no. 1 (1979). https://doi.org/10.2307/1599414.

Kelly, J. S., and B. S. Lindblom, eds. *Scientific Examination of Questioned Documents*. 2nd ed. Boca Raton: CRC Press—Taylor & Francis Group, 2006.

Kerkhoff, W., R. D. Stoel, C. E. Berger, E. J. Mattijssen, R. Hermsen, N. Smits, and H. J. Hardy. "Design and Results of an Exploratory Double Blind Testing Program in Firearms Examination." *Science & Justice: Journal of the Forensic Science Society* 55, no. 6 (Dec 2015): 514-9. https://doi.org/10.1016/j.scijus.2015.06.007.

Kerkhoff, W., R. D. Stoel, E.J.A.T. Mattijessen, R. Hermsen, P. Hertzman, D. Hazarda, M. Gallidabino, T. Hicks, and C. Champod. "Cartridge Case and Bullet Comparison: Examples of Evaluative Reporting." *Association of Firearm and Toolmark Examiners Journal* 49, no. 2 (2017): 111–21. https://nrl.northumbria.ac.uk/id/eprint/35853.

Kerstholt, J., A. Eikelboom, T. Dijkman, R. Stoel, R. Hermsen, and B. van Leuven. "Does Suggestive Information Cause a Confirmation Bias in Bullet Comparisons?". *Forensic Science International* 198, no. 1-3 (May 20 2010): 138-42. https://doi.org/10.1016/j.forsciint.2010.02.007.

Kirk, R. E. *Experimental Design.* 2nd ed. Belmont, CA: Wadsworth, Inc, 1982.

Koehler, J. "Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter." *Hastings Law Journal* 59, no. 5 (2008): 1077–100.

Koehler, J., and M. J. Saks. *Individualization Claims in Forensic Science: Still Unwarranted.* (2010). https://scholarlycommons.law.northwestern.edu/facultyworkingpapers/27.

Kolmogorov, A. N. *Foundations of the Theory of Probability.* New York: Chelsea Publishing Company, 1933.

Krane, D. E., S. Ford, J. R. Gilder, K. Inman, A. Jamieson, R. Koppl, I. L. Kornfield*, et al.* "Sequential Unmasking: a Means of Minimizing Observer Effects in Forensic DNA Interpretation." *Journal of Forensic Sciences* 53, no. 4 (Jul 2008): 1006-7. https://doi.org/10.1111/j.1556-4029.2008.00787.x.

Kukucka, J., and S. Kassin. "Do Confessions Taint Perceptions of Handwriting Evidence? An Empirical Test of the Forensic Confirmation Bias." *Law and Human Behavior* 38, no. 3 (2014): 256–70.

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

Langenburg, G., C. Champod, and P. Wertheim. "Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons." *Journal of Forensic Sciences* 54, no. 3 (May 2009): 571-82. https://doi.org/10.1111/j.1556-4029.2009.01025.x.

Lashley, K. S. "Mass Action in Cerebral Function." *Science* 73, no. 1888 (Mar 6 1931): 245-54. https://doi.org/10.1126/science.73.1888.245.

Leung, S. C., and Y. L.. Cheung. "On Opinion." *Forensic Science International* 42 (1989): 1–13.

Lindley, D. V. "A Problem in Forensic Science." *Biometrika* 64, no. 2 (1977): 207-13. https://doi.org/10.1093/biomet/64.2.207.

Lines, S., and F. E. Franck. "Triplet and Sibling Handwriting Study to Determine Degree of Individuality and Natural Variation." *Journal of the American Society of Questioned Document Examiners* 6 (2003): 48–55.

Liwicki, M., M. I. Malik, and C. E. H. Berger. "Towards a Shared Conceptualization for Automatic Signature Verification." In *Advances in Digital Handwritten Signature Processing*, edited by D. Impedovo G. Pirlo, and M. Fairhurst. 65–80. Singapore: World Scientif, 2014.

Liwicki, Marcus, Muhammad Imran Malik, C. Elisa van den Heuvel, Xiaohong Chen, Charles Berger, Reinoud Stoel, Michael Blumenstein, and Bryan Found. "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)." 2011 International Conference on Document Analysis and Recognition, 2011.

Malik, M. I., M. Liwicki, A. Dengel, and B. Found. "Man vs. Machine: A Comparative Analysis for Signature Verification." *Journal of Forensic Document Examination* 24 (2014): 21–35. https://doi.org/10.31974/jfde24-21-35.

Malik, Muhammad Imran, and Marcus Liwicki. "From Terminology to Evaluation: Performance Assessment of Automatic Signature Verification Systems." 2012 International Conference on Frontiers in Handwriting Recognition, Bari, Italy, IEEE, 2012.

Malik, Muhammad Imran, Marcus Liwicki, Linda Alewijnse, Wataru Ohyama, Michael Blumenstein, and Bryan Found. "ICDAR 2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp 2013)." 12th International Conference on Document Analysis and Recognition, 2013.

Marquis, R., A. Biedermann, L. Cadola, C. Champod, L. Gueissaz, G. Massonnet, W. D. Mazzella, F. Taroni, and T. Hicks. "Discussion on How to Implement a Verbal Scale in a Forensic Laboratory: Benefits, Pitfalls and Suggestions to Avoid Misunderstandings." *Science & Justice: Journal of the Forensic Science Society* 56, no. 5 (Sep 2016): 364-70. https://doi.org/10.1016/j.scijus.2016.05.009.

Marquis, R., S. Bozza, M. Schmittbuhl, and F. Taroni. "Handwriting Evidence Evaluation Based on the Shape of Characters: Application of Multivariate Likelihood Ratios." *Journal of Forensic Sciences* 56 Suppl 1 (Jan 2011): S238-42. https://doi.org/10.1111/j.1556-4029.2010.01602.x.

Martire, K. A., R. I. Kemp, I. Watkins, M. A. Sayle, and B. R. Newell. "The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect." *Law and Human Behavior* 37, no. 3 (Jun 2013): 197-207. https://doi.org/10.1037/lhb0000027.

Martire, K. A., and I. Watkins. "Perception Problems of the Verbal Scale: A Reanalysis and Application of a Membership Function Approach." *Science & Justice: Journal of the Forensic Science Society* 55, no. 4 (Jul 2015): 264-73. https://doi.org/10.1016/j.scijus.2015.01.002.

Mattijssen, E. J., W. Kerkhoff, C. E. Berger, I. E. Dror, and R. D. Stoel. "Implementing Context Information Management in Forensic Casework: Minimizing Contextual Bias in Firearms Examination." *Science & Justice: Journal of the Forensic Science Society* 56, no. 2 (Mar 2016): 113-22. https://doi.org/10.1016/j.scijus.2015.11.004. https://www.ncbi.nlm.nih.gov/pubmed/26976470.

McClelland, J., and D. Rumelhart. "An Interactive Activation Model of Context Effects in Letter Perception: Part 1, an Account of Basic Findings." *Psychological Review* 88, no. 2 (2011): 375-407.

McClure, D. *Focus Group on Scientific and Forensic Evidence in the Courtroom.* National Institute of Justice (Washington, DC: 2007). https://www.ncjrs.gov/pdffiles1/nij/grants/220692.pdf.

Meester, Ronald, and Marjan Sjerps. "Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence." *Law, Probability and Risk* 3, no. 1 (2004): 51-62. https://doi.org/10.1093/lpr/3.1.51.

Melendez-Diaz v. Massachusetts, 557 U.S. 305 (2009).

Melson, K. "Crime Laboratory Accreditation: The Assurance of Quality in the Forensic Sciences." In *The State of Criminal Justice*. American Bar Association, 2003.

———. "Improving the Forensic Sciences through Crime Laboratory Accreditation." In *Wiley Encyclopedia of Forensic Science*, Wiley-Blackwell, 2009.

Merlino, M. L., T. M. Freeman, V. Springer, V. Dahir, D. Hammond, A. D. Dyer, B. J. Found, L. Smith, and I. Duvall. *Validity, Reliability, Accuracy, and Bias in Forensic Signature Identification.* National Institute of Justice (2015). https://www.ncjrs.gov/pdffiles1/nij/grants/248565.pdf.

Merlino, M. L., C. I. Murray, and J. T. Richardson. "Judicial Gatekeeping and the Social Construction of the Admissibility of Expert Testimony." *Behavioral Sciences and the Law* 26, no. 2 (2008): 187-206. https://doi.org/10.1002/bsl.806.

Merlino, M. L., V. Springer, J. S. Kelly, D. Hammond, E. Sahota, and L. Haines. "Meeting the Challenges of the Daubert Trilogy: Refining and Redefining the Reliability of Forensic Evidence." *Tulsa Law Review* 43, no. 2 (2008): 417–46. https://heinonline.org/HOL/LandingPage?handle=hein.journals/tlj43&div=25&id=&page=.

Merlino, M. L., V. Springer, and A. Sigillo. "The Social Construction of the Admissibility of Most Frequently-Proffered Varieties of Expert Testimony." In *The Future of Evidence: How Science and Technology will Change the Practice of Law*, edited by C. Henderson and J. Epstein. 1–20. Chicago: American Bar Association, 2011.

Miller, J. J., R. B. Patterson, D. T. Gantz, C. P. Saunders, M. A. Walch, and J. Buscaglia. "A Set of Handwriting Features for Use in Automated Writer Identification." *Journal of Forensic Sciences* 62, no. 3 (May 2017): 722-34. https://doi.org/10.1111/1556-4029.13345. https://www.ncbi.nlm.nih.gov/pubmed/28054339.

Miller, Larry S. "Bias Among Forensic Document Examiners: A Need for Procedural Change." *Journal of Police Science & Administration* 12, no. 4 (1984): 407-11. https://psycnet.apa.org/record/1985-17219-001.

Miller v. State, 127 So.3d 580 (Fla. Dist. Ct. App. 2012).

Mnookin, J. L., S. A. Cole, I. E. Dror, B. A. J. Fisher, M. M. Houck, K. Inman, D. H. Kaye, *et al.* "The Need for a Research Culture in the Forensic Sciences." *UCLA Law Review* 58, no. 3 (2011): 725-80. https://heinonline.org/HOL/LandingPage?handle=hein.journals/uclalr58&div=20&id=&page=.

Mohammed, L., B. Found, and D. Rogers. "Frequency of Signature Styles in San Diego County." *Journal of the American Society of Questioned Document Examiners* 11, no. 1 (2008): 9–13.

Nakhaeizadeh, S., I. E. Dror, and R. Morgan. "The Emergence of Cognitive Bias in Forensic Science and Criminal Investigations." *British Journal of American Legal Studies* 4 (2015): 527–54.

Nakhaeizadeh, S., I. E. Dror, and R. M. Morgan. "Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias." *Science & Justice: Journal of the Forensic Science Society* 54, no. 3 (May 2014): 208-14. https://doi.org/10.1016/j.scijus.2013.11.003.

National Association of Testing Australia (NATA). *General Accreditation Criteria: Legal Management of Facility Activities (Forensic Operations Module).* (2019). https://www.nata.com.au/phocadownload/gen-accreditation-criteria/Forensic-Operations-Module.pdf.

———. *Specific Accreditation Criteria: ISO/IEC 17025 Application Document Legal (including Forensic Science)—Appendix.* (2018). https://www.nata.com.au/phocadownload/spec-criteria-guidance/legal-forensic/Forensic-Science-ISO-IEC-17025-Appendix.pdf.

National Association of Testing Authorities (NATA), Australia. "National Association of Testing Authorities (NATA), Australia home page." 2019, accessed May 6, 2020, https://www.nata.com.au/.

National Commission on Forensic Science (NCFS). *Directive Recommendation: Root Cause Analysis (RCA) in Forensic Science.* Department of Justice (2015). https://www.justice.gov/archives/ncfs/file/786581/download.

———. *Proficiency Testing in Forensic Science.* Department of Justice (2016). https://www.justice.gov/archives/ncfs/page/file/831806/download.

———. *Recommendation to the Attorney General Regarding Transparency of Quality Management System Documents.* Department of Justice (2016). https://www.justice.gov/ncfs/file/839706/download.

———. *Recommendation to the Attorney General: National Code of Professional Responsibility for Forensic Science and Forensic Medicine Service Providers.* Department of Justice (2016). https://www.justice.gov/ncfs/file/839711/download.

———. *Recommendation to the Attorney General: Universal Accreditation.* Department of Justice (2015). https://www.justice.gov/archives/ncfs/file/477851/download.

———. *Recommendations to the Attorney General: Pretrial Discovery.* Department of Justice (2016). https://www.justice.gov/ncfs/file/880241/download.

———. *Views of the Commission: Documentation, Case Record and Report Contents.* Department of Justice (2015). https://www.justice.gov/ncfs/file/818191/download.

———. *Views of the Commission: Ensuring that Forensic Analysis Is Based Upon Task-Relevant Information.* Department of Justice (2015). https://www.justice.gov/archives/ncfs/file/818196/download.

———. *Views of the Commission: Facilitating Research on Laboratory Performance.* Department of Justice (2016). https://www.justice.gov/archives/ncfs/page/file/909311/download.

———. *Views of the Commission: Pretrial Discovery of Forensic Materials.* Department of Justice (2015). https://www.justice.gov/ncfs/file/786611/download.

National Institute of Standards and Technology (NIST). *Forensic Science Laboratories: Handbook for Facility Planning, Design, Construction, and Relocation.* U.S. Department of Commerce (June 2013 2013).

———. *Handbook (NIST HB) 150.* 2006.

National Research Council (NRC). *DNA Technology in Forensic Science.* The National Academies (Washington, DC: 1992).

———. *Strengthening Forensic Science in the United States: A Path Forward.* The National Academies Press (Washington, DC: 2009).

National Science and Technology Council, Subcommittee on Forensic Science. *Strengthening the Forensic Sciences.* National Science and Technology Council (Washington, DC: 2014). https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/forensic_science___may_2014.pdf.

National Science Foundation (NSF). "Transdisciplinary Research in Principles of Data Science (TRIPODS)." Updated March 19, 2019, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505347.

Neumann, C., and V. Reyna. *Jury Studies and the Psychology of Efficient Communication.* NIST (11/8/15 2015). https://www.researchgate.net/publication/291817447_Communicating_the_Results_of_Forensic_Science_Examinations.

New York City Office of Chief Medical Examiner. *Forensic Biology Evidence and Case Management Manual.* (2019). https://www1.nyc.gov/assets/ocme/downloads/pdf/technical-manuals/forensic-biology-evidence-and-case-management-manual/administrative-review.pdf.

Nickerson, Raymond S. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2, no. 2 (1998): 175-220. https://doi.org/10.1037/1089-2680.2.2.175.

Nordgaard, A., R. Ansell, W. Drotz, and L. Jaeger. "Scale of Conclusions for the Value of Evidence." *Law, Probability and Risk* 11, no. 1 (2011): 1-24. https://doi.org/10.1093/lpr/mgr020.

North Carolina State Crime Laboratory. *Procedure for Reviewing Laboratory Reports.* (2013). https://forensicresources.org/wp-content/uploads/2019/07/Reviewing-Laboratory-Reports-08-29-2014.pdf.

Nunnally, J. C., and I. H. Bernstein. *Psychometric Theory.* 3rd ed. New York: McGraw-Hill Publishing Co., 1994.

Office of Management and Budget. *Final Information Quality Bulletin for Peer Review.* (December 15 2004). https://www.epa.gov/sites/production/files/2015-01/documents/omb_final_info_quality_bulletin_peer_review_2004_1.pdf.

Organization of Scientific Area Committees (OSAC) for Forensic Science. *OSAC Standards Bulletin.* (2017). https://www.nist.gov/system/files/documents/2017/10/24/osac_standards_bulletin_october_2017_2.pdf.

Orne, Martin T. "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications." *American Psychologist* 17, no. 11 (1962): 776-83. https://doi.org/10.1037/h0043424.

Osborn, A. S. *The Problem of Proof: Especially as Exemplified in Disputed Document Trials: A Discussion of the Proof of the Facts in Courts of Law: With Some General Comments on the Conduct of Trials.* Newark, NJ: Essex Press, 1926.

———. *Questioned Documents.* Second ed. Albany, NY: Boyd Printing Company, 1929.

Osborne, N. K. P., and M. C. Taylor. "Contextual Information Management: An Example of Independent-Checking in the Review of Laboratory-Based Bloodstain Pattern Analysis." *Science & Justice: Journal of the Forensic Science Society* 58, no. 3 (May 2018): 226-31. https://doi.org/10.1016/j.scijus.2018.01.001.

Osborne, N. K., M. C. Taylor, M. Healey, and R. Zajac. "Bloodstain Pattern Classification: Accuracy, Effect of Contextual Information and the Role of Analyst Characteristics." *Sci Justice* 56, no. 2 (Mar 2016): 123-8. https://doi.org/10.1016/j.scijus.2015.12.005.

Osborne, N. K., S. Woods, J. Kieser, and R. Zajac. "Does Contextual Information Bias Bitemark Comparisons?". *Science & Justice: Journal of the Forensic Science Society* 54, no. 4 (Jul 2014): 267-73. https://doi.org/10.1016/j.scijus.2013.12.005.

Oswald, M. E., and S. Grosjean. "Confirmation Bias." Chap. 4 In *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*, edited by R. F. Pohl. 79–96. Hove and NY: Psychology Press, 2004.

Page, M., J. Taylor, and M. Blenkin. "Forensic Identification Science Evidence Since Daubert: Part II—Judicial Reasoning in Decisions to Exclude Forensic Identification Evidence on Grounds of Reliability." *J Forensic Sci* 56, no. 4 (Jul 2011): 913-7. https://doi.org/10.1111/j.1556-4029.2011.01776.x.

———. "Uniqueness in the Forensic Identification Sciences—Fact or Fiction?". *Forensic Science International* 206, no. 1-3 (Mar 20 2011): 12-8. https://doi.org/10.1016/j.forsciint.2010.08.004.

Parker, J. B. "A Statistical Treatment of Identification Problems." *Journal of the Forensic Science Society* 6, no. 1 (1966): 33–39.

Paulhus, Delroy L. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes*, edited by P. R. Shaver and L. S. Wrightsman J. P. Robinson. 17-59. San Diego, CA: Academic Press, 1991.

Pettus v. United States, 37 A.3d 213 (D.C. 2012).

Plamondon, Réjean, and Guy Lorette. "Automatic Signature Verification and Writer Identification—the State of the Art." *Pattern Recognition* 22, no. 2 (1989): 107-31. https://doi.org/10.1016/0031-3203(89)90059-9.

Preparing Future Faculty (PFF). "About PFCF/PFF." University of Kentucky, n.d., accessed May 11, 2020, https://www.uky.edu/CommInfoStudies/GRAD/PFF/about.html.

President's Council of Advisors on Science and Technology (PCAST). *An Addendum to the PCAST Report on Forensic Science in Criminal Courts.* (January 6, 2017 2017). https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf.

———. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods.* President's Council of Advisors on Science and Technology (Washington, DC: 2016). https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

Quattrone Center for the Fair Administration of Justice. *Guidelines for the Use of Root Cause Analysis (RCA) to Reduce Error and Improve Quality in Forensic Science Laboratories.* (2014).

https://www.nist.gov/system/files/documents/2016/11/22/guidelines_for_the_use_of_root_cause_analysis_to_reduce_error_and_improve_quality_in_forensic_science_labs.hollway.labmgmt.pdf.

Queensland Police Service. *Forensic Services Quality Manual.* (2015).

Reason, J. "Human Error: Models and Management." *BMJ* 320, no. 7237 (Mar 18 2000): 768-70. https://doi.org/10.1136/bmj.320.7237.768.

Risinger, D. M. "The NAS Report on Forensic Science: a Glass Nine-Tenths Full (This is About the Other Tenth)." *Jurimetrics* 50, no. 1 (2009): 21–34.

Risinger, D. M., M. J. Saks, W. C. Thompson, and R. Rosenthal. "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion." *California Law Review* 90, no. 1 (2002): 1–56. https://heinonline.org/HOL/LandingPage?handle=hein.journals/calr90&div=10&id=&page=.

Robertson, B., G. A. Vignaux, and C. E. H. Berger. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom.* 2nd ed. Chichester: Wiley, 2016.

Robertson, C. T., and A. S. Kesselheim, eds. *Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law*. Atlanta, GA: Elsevier, 2016.

Royall, R. *Statistical Evidence: A Likelihood Paradigm.* Monographs on Statistics and Applied Probability. Chapman & Hall/CRC Press LLC, 1997.

Said, H. E. S., T. N. Tan, and K. D. Baker. "Personal Identification Based on Handwriting." *Pattern Recognition* 33, no. 1 (2000): 149-60. https://doi.org/10.1016/s0031-3203(99)00006-0.

Saks, M.J., and J. J. Koehler. "The Individualization Fallacy in Forensic Science Evidence." *Vanderbilt Law Review* 61, no. 1 (2008): 199–219.

Saunders, C. P., L. J. Davis, and J. Buscaglia. "Using Automated Comparisons to Quantify Handwriting Individuality." *Journal of Forensic Sciences* 56, no. 3 (May 2011): 683-9. https://doi.org/10.1111/j.1556-4029.2011.01713.x.

Saunders, C. P., L. J. Davis, A. C. Lamas, J. J. Miller, and D. T. Gantz. "Construction and Evaluation of Classifiers for Forensic Document Analysis." *Annals of Applied Statistics* 5, no. 1 (2011): 381–99.

Schiff, G. D. "Minimizing Diagnostic Error: the Importance of Follow-Up and Feedback." *The American Journal of Medicine* 121, no. 5 Suppl (May 2008): S38-42. https://doi.org/10.1016/j.amjmed.2008.02.004.

Schiffer, B., and C. Champod. "The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation." *Forensic Science International* 167, no. 2-3 (Apr 11 2007): 116-20. https://doi.org/10.1016/j.forsciint.2006.06.036.

Schomaker, L. "Writer Identification and Verification." In *Advances in Biometrics*, edited by N. K. Ratha and V. Govindaraju. 247–64. London: Springer, 2008.

Schuck, Julie Anne. *Personnel Selection in the Pattern Evidence Domain of Forensic Science.* 2017. doi:10.17226/23681.

Schweitzer, N. J. *Communicating Forensic Science.* National Criminal Justice Reference Service (2016). https://www.ncjrs.gov/pdffiles1/nij/grants/249804.pdf.

Scientific Working Group for Forensic Document Examination (SWGDOC). "Guidelines for Forensic Document Examination." *FBI Forensic Science Communications* 2, no. 2 (2000): 1-8.

———. "Scientific Working Group for Forensic Document Examination (SWGDOC) home page." Updated March 20, 2018, 2019, https://www.swgdoc.org/.

———. *SWGDOC Standard for Examination of Handwritten Items.* Version 2013-1. 2013.

———. *SWGDOC Standard for Minimum Training Requirements for Forensic Document Examiners.* Version 2013-1. 2013.

———. *SWGDOC Standard for Use of Image Capture and Storage Technology in Forensic Document Examination.* Version 2013-1. 2013.

———. *SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners.* Version 2013-2. 2013.

Scientific Working Group on Friction Ridge Analysis, Study and Technology. *A Model Policy for Friction Ridge Examiner Professional Conduct.* Scientific Working Group on Friction Ridge Analysis, Study and Technology (2008).

Shafer, Glenn. "Lindley's Paradox." *Journal of the American Statistical Association* 77, no. 378 (1982). https://doi.org/10.1080/01621459.1982.10477809.

Shappell, S. A., and D. A. Wiegmann. *The Human Factors Analysis and Classification System–HFACS.* Office of Aviation Medicine (Washington, DC: 2000). https://www.nifc.gov/fireInfo/fireInfo_documents/humanfactors_classAnly.pdf.

Sieden, H., and F. Norwitch. "Questioned Documents." In *Forensic Science: An Introduction to Scientific and Investigative Techniques*, edited by S. H. James, J. J. Norby, and S. Bell. Boca Raton: CRC Press, 2014.

Siegel, Jay A., Megan King, and Whitney Reed. "The Laboratory Report Project." *Forensic Science Policy & Management: An International Journal* 4, no. 3-4 (2013): 68-78. https://doi.org/10.1080/19409044.2013.858798.

Sita, Jodi, Bryan Found, and Douglas K. Rogers. "Forensic Handwriting Examiners' Expertise for Signature Comparison." *Journal of Forensic Sciences* 47, no. 5 (2002). https://doi.org/10.1520/jfs15521j.

Smith v. State, 702 N.E.2d 668, 673 (Ind. 1998).

Song, J. "Proposed "NIST Ballistics Identification System (NBIS)" Based on 3D Topography Measurements on Correlation Cells." *AFTE Journal* 45, no. 2 (2013): 184-94. https://pdfs.semanticscholar.org/1103/38f65f80958b94317bfd780556231adf0995.pdf.

Srihari, S., C. Huang, and H. Srinivasan. "On the Discriminability of the Handwriting of Twins." *Journal of Forensic Sciences* 53, no. 2 (Mar 2008): 430-46. https://doi.org/10.1111/j.1556-4029.2008.00682.x.

Srihari, S.N. *Computational Methods for Handwritten Questioned Document Examination.* (2010). https://www.ncjrs.gov/pdffiles1/nij/grants/232745.pdf.

Srihari, S.N., S. Cha, H. Arora, and S. Lee. *Individuality of Handwriting.* (2001). https://www.ncjrs.gov/pdffiles1/nij/grants/190133.pdf.

Srihari, Sargur N., Sung-Hyuk Cha, Hina Arora, and Sangjik Lee. "Individuality of Handwriting." *Journal of Forensic Sciences* 47, no. 4 (2002). https://doi.org/10.1520/jfs15447j.

Stamatis, D. H. *Failure Mode and Effect Analysis: FMEA from Theory to Execution.* 2nd ed.: ASQ Quality Press, 2003.

State v. Reid, 757 A.2d 482, 487 (Conn. 2000).

State v. Small, 184 A.3d 816, 825 (Conn.App. 2018).

Stevens, S. S. "On the Theory of Scales of Measurement." *Science* 103, no. 2684 (Jun 7 1946): 677-80. https://doi.org/10.1126/science.103.2684.677.

Stoel, R. D., C. E. H. Berger, W. Kerkhoff, E. J. A. T. Mattijssen, and I. E. Dror. "Minimizing Contextual Bias in Forensic Casework." In *Forensic Science and the Administration of Justice: Critical Issues and Directions*, edited by Kevin Strom and Matthew Hickman. Thousand Oaks: SAGE Publications, Inc, 2014.

Stoel, R. D., W. Kerkhoff, Ejat Mattijssen, and C. E. H. Berger. "Building the Research Culture in the Forensic Sciences: Announcement of a Double Blind Testing Program." *Science & Justice: Journal of the Forensic Science Society* 56, no. 3 (May 2016): 155-56. https://doi.org/10.1016/j.scijus.2016.04.003.

Stoney, D. A. "Evaluation of Associative Evidence: Choosing the Relevant Question." *Journal of the Forensic Science Society* 24, no. 5 (1984): 473-82. https://doi.org/10.1016/s0015-7368(84)72326-7.

Taupin, J. M. *Introduction to Forensic DNA Evidence for Criminal Justice Professionals.* Boca Raton: CRC Press, 2013.

Tensor Flow. "Tensor Flow home page." 2019, accessed May 6, 2020, https://www.tensorflow.org/.

The American Academy of Forensic Sciences (AAFS). "Types of Forensic Scientists: Disciplines of AAFS." 2019, accessed May 6, 2020, https://www.aafs.org/aafs/Resources/Students/Choosing-a-Career/Types-of-Forensic-Scientist--Disciplines-of-AAFS/AAFS/Resources/Students/Types.aspx?hkey=a401388a-32dc-4c74-aee5-c52de769c6e0.

The Expert Working Group on Human Factors in Latent Print Analysis. *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach.* U.S. Department of Commerce, National Institute of Standards and Technology (2012).

The Innocence Project. "Overturning Wrongful Convictions Involving Misapplied Forensics." n.d., accessed May 6, 2020, https://www.innocenceproject.org/causes/misapplication-forensic-science/.

The International Review of Research in Open and Distributed Learning. "The International Review of Research in Open and Distributed Learning home page." 2019, accessed May 6, 2020, www.irrodl.org.

Thompson, W. C. "Painting the Target Around the Matching Profile: the Texas Sharpshooter Fallacy in Forensic DNA Interpretation." *Law, Probability and Risk* 8, no. 3 (2009): 257-76. https://doi.org/10.1093/lpr/mgp013.

Thompson, William C. "What Role Should Investigative Facts Play in the Evaluation of Scientific Evidence?". *Australian Journal of Forensic Sciences* 43, no. 2-3 (2011): 123-34. https://doi.org/10.1080/00450618.2010.541499.

Thompson, William C., Suzanne O. Kaasa, and Tiamoyo Peterson. "Do Jurors Give Appropriate Weight to Forensic Identification Evidence?". *Journal of Empirical Legal Studies* 10, no. 2 (2013): 359-97. https://doi.org/10.1111/jels.12013.

Thompson, William C., and Edward L. Schumann. "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy." *Law and Human Behavior* 11, no. 3 (1987): 167-87. https://doi.org/10.1007/bf01044641.

Trinder, J. C. "Competency Standards—a Measure of the Quality of a Workforce." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVII, no. Part B6a (2008): 165–68.

Triplett, M., and L. Cooney. "Etiology of ACE-V and Its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing." *Journal of Forensic Identification* 56, no. 3 (2006): 345–55. https://pdfs.semanticscholar.org/0b71/cba100dce3da6266d22df6045c37faf4ee77.pdf.

Tversky, Amos, and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5, no. 2 (1973): 207-32. https://doi.org/10.1016/0010-0285(73)90033-9.

U.S. Army Criminal Investigation Laboratory. *Program of Instruction for Document Examination Course.* (October 1966).

U.S. Department of Justice. *Code of Professional Responsibility for the Practice of Forensic Science.* (2016). https://www.justice.gov/sites/default/files/code_of_professional_responsibility_for-the_practice_of_forensic_science_08242016.pdf.

———. "Justice Department Announces New Accreditation Policies to Advance Forensic Science." News release, December 7, 2015, https://www.justice.gov/opa/pr/justice-department-announces-new-accreditation-policies-advance-forensic-science.

U.S. Department of Justice, Office of Justice Programs. *Education and Training in Forensic Science: A Guide for Forensic Science Laboratories, Educational Institutions, and*

Students. Technical Working Group for Education and Training in Forensic Science (TWGED) (April 2004).

———. *Forensic Sciences: Review of Status and Needs.* (February 1999).

U.S. Department of Justice, Office of the Inspector General,. *The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases.* Office of the Inspector General (Washington, DC: April 1997).

U.S. Department of Labor Bureau of Labor Statistics. "Labor Force Statistics from the Current Population Survey." Updated February 8, 2017, https://www.bls.gov/cps/cpsaat18b.htm.

Ulery, B. T., R. A. Hicklin, J. Buscaglia, and M. A. Roberts. "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners." *PLoS One* 7, no. 3 (2012): e32800. https://doi.org/10.1371/journal.pone.0032800.

United States v. Anderson, 169 F.Supp.3d 60 (D.D.C. 2016).

United States v. Brewer, No. 01 CR 892, 2002 U.S. Dist. LEXIS 6689 (N.D.Ill 2002).

United States v. Chischilly, 30 F.3d 1144 (9th Cir. 2014).

United States v. Davis, 514 F.3d 596 (6th Cir. 2008).

United States v. Fujii, 152 F. Supp. 2d 939 (N.D.Ill. 2000).

United States v. Gricco, No. 01-90, 2002 WL 746037, 2002 U.S. Dist. LEXIS 7564 (E.D.Pa. 2002).

United States v. Hidalgo, 229 F. Supp. 2d 961, 968 (D. Ariz. 2002).

United States v. Hines, 55 F. Supp. 2d 62 (D.Mass. 1999).

United States v. Janati, 374 F.3d 263, 273 (4th Cir. 2004).

United States v. Johnsted, 30 F. Supp. 3d 814, 821 (W.D. Wis. 2013).

United States v. Lewis, 220 F. Supp. 2d 548 (S.D.W. Va. 2002).

United States v. Prime, 220 F. Supp. 2d 1203 (W.D. Wash. 2002).

United States v. Richmond, No. 00-321,, 2001 WL 1117235, 2001 U.S. Dist. LEXIS 15769 (E.D.La. 2001).

United States v. Rutherford, 104 F. Supp. 2d 1190 (D. Neb. 2000).

United States v. Saelee, 162 F. Supp. 2d 1097, 1103 (D. Alaska 2001).

United States v. Santillan, No. CR-96-40169, 1999 U.S. Dist. LEXIS 21611 (N.D.Cal. 1999).

United States v. Shue, 766 F.2d 1122 (7th Cir. 1985).

United States v. Smith, 101 F.3d 202 (1st Cir. 1996).

United States v. Walker, 910 F.Supp. 861 (N.D.N.Y. 1995).

United States v. Williams, 583 F.2d 1194, 1198 (2nd Cir. 1978).

van Es, A., W. Wiarda, M. Hordijk, I. Alberink, and P. Vergeer. "Implementation and Assessment of a Likelihood Ratio Approach for the Evaluation of LA-ICP-MS Evidence in Forensic Glass Analysis." *Science & Justice: Journal of the Forensic Science Society* 57, no. 3 (May 2017): 181-92. https://doi.org/10.1016/j.scijus.2017.03.002.

Venville, Nikkita. *A Review of Contextual Bias in Forensic Science and Its Potential Legal Implications. Report of the Victorian Law Foundation Legal Policy Internship Program.* The Victoria Law Foundation, The Australia New Zealand Police Advisory Authority and the National Institute of Forensic Science (ANZPAA-NIFS) (2011). https://netk.net.au/Psychology/Psychology14.pdf.

Verma, E. "From Traditional Classroom to Online Learning: ICT as Change Agent to Boost Innovation in Teaching and Learning." *International Journal of Scientific Research in Science and Technology* 3, no. 8 (2017): 155–60.

Virginia Department of Forensic Science. *Quality Manual.* (2017). https://www.dfs.virginia.gov/wp-content/uploads/2020/09/100-D100-DFS-Quality-Manual.pdf.

Wagenaar, W. A. *Identifying Ivan: A Case Study in Legal Psychology.* London: Harvester/Wheatsheaf, 1988.

Williams v. Illinois, 567 U.S. 50 (2012).

Wilson-Wilde, Linzi, Stephen Smith, and Eva Bruenisholz. "The Analysis of Australian Proficiency Test Data over a Ten-Year Period." *Forensic Science Policy & Management: An International Journal* 8, no. 1-2 (2017): 55-63. https://doi.org/10.1080/19409044.2017.1352054.

Wilson, T. D., and N. Brekke. "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations." *Psychological Bulletin* 116, no. 1 (Jul 1994): 117-42. https://doi.org/10.1037/0033-2909.116.1.117.

Wing, Alan M. "Motor Control: Mechanisms of Motor Equivalence in Handwriting." *Current Biology* 10, no. 6 (2000): R245-R48. https://doi.org/10.1016/s0960-9822(00)00375-4.

Zheng, R., Y. Qin, Z. Huang, and H. Chen. "Authorship Analysis in Cybercrime Investigation." In *Intelligence and Security Informatics*, edited by H. Chen, R. Miranda, D. D. Zeng, C. Demchak, J. Schroeder and T. Madhusudan. 59-73. Berlin, Heidelberg: Springer, 2003.