



J1 The Accuracy and Reliability of Forensic Handwriting Examiner Decisions

Linda Eisenhart*, Federal Bureau of Investigation, Quantico, VA 22135; R. Austin Hicklin, PhD*, Noblis, Reston, VA 20191; Brian Eckenrode, PhD, Quantico, VA 22135; Peter J. Belcastro, Jr., MA, Federal Bureau of Investigation, Quantico, VA 22135; Ted M. Burkes, BS, Stafford, VA 22554; Meredith DeKalb Miller, MFS, Davenport, IA 52803; Michael A. Smith, PhD, Federal Bureau of Investigation Laboratory, Quantico, VA 22135; Connie L. Parks, MA, Washington, DC 20024; JoAnn Buscaglia, PhD, Federal Bureau of Investigation Laboratory, CFSRU, Quantico, VA 22135

Learning Overview: After attending this presentation, attendees will have learned about the results of the large-scale Handwriting Decision Analysis (Black Box) Study, including accuracy, error rate(s), reproducibility, and repeatability of handwriting decisions, and what factors impact these rates.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by enabling forensic document examiners to answer questions regarding what factors impact the accuracy and reproducibility of the decisions, including those related to the examiner and those related to the samples.

This Handwriting Decision Analysis (also known as Black Box) Study was designed and conducted by the Federal Bureau of Investigation (FBI) Laboratory, Ideal Innovations, and Noblis. The design of this study was based on that of the FBI Laboratory-Noblis latent print examiner black box study in 2011, and in consideration of comments made in two United States Reports: the 2009 National Academy of Sciences Report, *Strengthening Forensic Science in the United States: A Pathway Forward*, and the 2016 Report from the President's Council of Advisors on Science & Technology titled *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*.¹⁻³

Participation for this study was limited to practicing handwriting examiners who had performed casework within the past two years. A pre-test survey was completed by each participant to gather information about their education, training, experience, certification, and day-to-day work. Participants were asked to perform a total of 100 comparisons of paired handwriting samples, each containing one sample of questioned writing and up to five pages of writing from a single known subject. In an effort to represent the wide range of evidence encountered in casework, a diversity of handwriting samples was selected. The test samples included a range of styles of handwriting (i.e., cursive, hand printing, and mixed), content (e.g., addresses, a few sentences, predefined London or Center of Excellence for Document Analysis and Recognition (CEDAR) letters, or freeform text), known writing samples containing the same or different wording as well as varying amounts of questioned and known writing. Note that this study was limited to the comparison of handwritten documents and did not include the comparisons of signatures. Great effort was taken to provide participants with test packets that were relatively challenging, both for mated and non-mated pairs.

Using an online platform, participants were provided with high-resolution digital images that could be downloaded for comparison. The participants used this platform to respond to predefined questions that included the conclusion, limitation(s), assessments of comparability and variation within questioned and known writing samples, perceived difficulty of the comparison, and most influential characteristics that were used in making their decisions. Participants were given five conclusions to choose from, including definitive opinions (written by and not written by), qualified opinions (probably written by and probably not written by), and no conclusion. The study resulted in 7,213 trials from 86 handwriting examiners on 180 distinct comparisons of questioned and known handwriting.

This presentation will provide the results of the study, including the accuracy, reproducibility (inter-examiner variation), and repeatability (intra-examiner variation) of decisions made by forensic handwriting examiners. Also presented will be the factors that did (or did not) impact the accuracy of the decisions in this study, including those related to the examiner (such as training or experience) and those related to the samples (such as quantity of writing, comparability of content, limitations, or style of writing).

Reference(s):

1. Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proceeding of the National Academy of Sciences of the United States of America (PNAS)*. First published April 25, 2011 <https://doi.org/10.1073/pnas.1018707108>.
2. National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press, 2009.
3. President's Council of Advisors on Science and Technology (PCAST). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods*. Washington, DC: Office of Science and Technology, 2016.

Handwriting, Error Rates, Black Box Study