Technical Note

A Method for Characterizing Questioned Footwear Impression Quality

Brian C. McVicker [1]
Connie Parks [2]
Jan LeMay [3]
Brian A. Eckenrode [1]
R. Austin Hicklin [2]

Abstract: The quality of a questioned footwear impression affects a forensic footwear examiner's ability to conduct an accurate and useful footwear evidence examination. Here we introduce a reproducible and quantifiable framework specifically developed to assess the quality of questioned footwear impressions. We propose that this method holds the potential to be the basis for a standard system for the characterization and quantification of the quality of questioned footwear impressions.

Introduction

The forensic footwear examiner's (FFE's) ability to conduct an accurate and useful footwear evidence examination is affected by the quality of the evidence presented for examination. The National Research Council of the National Academies discussed the effects of varying sample quality, stating with respect to multiple forensic disciplines that the value of evidence is determined by the quality of the images or other samples [1]. The 2016 President's Council of Advisors on Science and Technology's critique of certain forensic disciplines, particularly pattern

[1] FBI Laboratory, Quantico, VA
[2] Noblis, Reston, VA
[3] Denver Police Department Crime Lab, Denver, CO

Received February 17, 2021; accepted March 29, 2021

evidence disciplines, contends that scientific studies intended to estimate a method's accuracy, reproducibility, and repeatability must be based on samples that reflect the distribution of characteristics found in casework [2]. When selecting data for validity studies, a metric for assessing the quality and quantity of information in an evidence impression is necessary in order to understand the distribution of sample characteristics and to ensure consistency in the quality of samples presented to examiners.

Data collection for an FFE decision analysis study (hereinafter "FFE Black Box Study") sponsored by the Federal Bureau of Investigation (FBI) Laboratory Division was completed in 2020. The authors were part of the research team who executed that study, the objectives of which were to estimate the accuracy, reproducibility, and repeatability of FFE decisions and to assess validity as applied. Because previous validation studies in latent fingerprint examination reported associations between examiner conclusions and sample quality (e.g., [3–5]), during the design phase of that study, we hypothesized that the quality of the questioned footwear impressions provided to study participants would affect their conclusions. Currently, there are no standardized quality metrics that we could rely upon to grade questioned footwear impressions. The *SWGTREAD Guide for the Examination of Footwear and Tire Impression Evidence* [6] does not provide sufficient detail regarding assessing a questioned impression's quality and suitability for comparison, nor does it provide a standard method for assessing quality or suitability. Other footwear impression quality assessment approaches have been reported [7, 8], but they were not sufficiently detailed for our study. Therefore, we developed a novel framework to assess and rate the quality of questioned footwear impressions. This footwear-specific rubric provided a method of assessing quality, which enabled us to construct a sample set approximating the wide range of quality levels found in evidential samples.

The results of the FFE Black Box Study (which have not yet been reported) demonstrate that there is an association between the quality of the footwear impressions as measured using this method, and the accuracy, reproducibility, and repeatability of study participants' conclusions. We therefore believe that this framework can provide the foundation for future discipline-specific quality assessment methods for use in both research and operations.

Defining Footwear Impression Quality

Introduced here is the method developed and employed by the study team to rate the quality of questioned impressions (or derivatives, such as images, lifts, and casts) created during the FFE Black Box Study. This method was developed through the team's discussions during which the FFEs on the team were asked to articulate what constitutes a high-quality impression. The term quality may be used to convey a variety of meanings in different forensic disciplines. The quality of a sample as defined by the International Organization for Standardization (ISO) [9] encompasses character, fidelity, and utility. Here we build on the ISO definition to be specific to footwear evidence:

- the character of a footwear impression refers to the quantity (or relative amount) of outsole features on the source footwear item that are reproduced in the impression, and the specificity (distinctiveness) of those features;
- the fidelity of a footwear impression refers to the accuracy with which the outsole features on the source footwear item are reproduced in the impression;
- the utility of a footwear impression refers to the impression's suitability (or value) for comparison and source attribution.

Note that utility is a function of character and fidelity. For example, a high-fidelity impression from a new footwear item is of limited utility because it is a low-character impression: the impression accurately reproduces class characteristics but cannot be discriminated from others of the same class given its absence of randomly acquired characteristics (RACs)[1]. Conversely, an impression that contains class characteristics and many RACs (high character) would be of limited utility if it is highly distorted (low fidelity). A small impression may be of high utility if the impression clearly reproduces highly distinctive features, such as a clear (high fidelity) impression of the toe area of an outsole with many RACs (high character that is due to the specificity of the features).

[1] A "randomly acquired characteristic" is a feature (e.g., a cut, a scratch, a tear, a hole, or a stone hold) on the outsole of a footwear item acquired through random events. The position, orientation, size, and shape of these characteristics can be used to differentiate one footwear outsole from another when those outsoles share the same class characteristics.

Our team characterized the highest quality questioned impressions as the ones that mimic the level of quality of known footwear test impressions prepared under controlled laboratory conditions. Based on this characterization, we developed 10 attributes as subcategories of character, fidelity, and utility, as shown in Table 1. The team developed considerations for each attribute in the form of a statement or question and three possible responses for each consideration which were linked to a 0 (poor) to 2 (good) score. The overall quality assessment is based on the sum of these; an ideal impression would receive a 2 in each attribute and therefore a 20 overall. The resulting scale is an ordinal scale that is monotonic with respect to quality. In other words, any higher score indicates higher quality than any lower score, but an overall score of 20 does not indicate twice the quality of an overall score of 10. The overall quality assessment is intended for ranking impressions by quality. The attributes are interrelated and are not intended to be independent. For example, distortion, overlap, substrate, and matrix all affect clarity, which affects pattern, which affects suitability. This rubric provided a standard way for the team's FFEs to assess each attribute for each impression with some level of consistency, and we believe that this may serve as a basis for a more widely used quality assessment metric. If the rubric is adopted for formal standardization, then we assume that the language and descriptions presented here may be enhanced as part of that process.

ISO	Attribute	Consideration	Assessment Score		
			0	1	2
Character	Quantity	Estimate the relative amount of the outsole that is reproduced in the impression.	Much less than half	About half	Most or all (heel to toe)
Fidelity	Pattern	Can you discern the geometric shapes that form the pattern in the impression?	No	Somewhat	Yes
	Contrast	Rate the contrast between the impression and the background.	Poor	Moderate	Good
	Distortion	How much distortion is present in the impression?	Significant amount	Some	None
	Substrate	Do features of the substrate (e.g., texture, voids, and background pattern) interfere with visualizing the impression detail?	Yes	Somewhat	No
	Matrix	Does the amount or type of matrix (e.g., too much or too little blood) prevent visualizing the impression detail?	Yes	Somewhat	No
	Overlap	Can you distinguish the primary impression from the overlapping impression(s)?	Hard to distinguiuish	Easy to distinguish	No overlapping impressions
	Clarity	Is the clarity of the impression sufficient to visualize fine detail (e.g., outsole texturing and potential RACs)?	No	Only in some areas	Yes
Character/ Fidelity	Left vs Right	Can you determine if the impression was made by a left or a right shoe?	No	Possibly, but uncertain	Yes, easy to determine
Utility	Suitability	Classify the impression according to its expected suitability for comparison.	Unsuitable for comparison	Suitable for class inclusion or exclusion	Suitable for identification

Table 1

Footwear impression quality assessment rubric as implemented in the FFE Black Box Study.

For the FFE Black Box Study, 253 impressions from 239 footwear items were collected and assessed using this quality rubric (some footwear items were used to create multiple impressions before and after wear). However, not all those impressions were used in that study, including the five impressions depicted in Figures 1 through 5. The five impressions illustrate how the quality rubric applies to a range of footwear impressions (Tables 2 through 6). In these images, lifts are reversed (i.e., the images were flipped horizontally) so that they orient with the impression on the ground. At least two FFEs (out of a team of four FFEs) independently assessed each impression, provided as high-resolution digital images viewed on-screen. When multiple images were available, the quality assessment was based on the totality of the information contained in all reproductions of an impression, as is the practice in casework. For each attribute, the average of the FFEs' assessments were used. Out of 2,530 attributes (10 each for 253 impressions), the FFEs agreed exactly on 67.7% of the assessments, and they completely disagreed (i.e., attribute assessments of 0 vs. 2) on 2.6% of the assessments. When the quality assessments from the initial two FFEs were notably different, the impression was assigned to a third (and sometimes fourth) FFE; quality assessments were assigned to additional FFEs when the total quality differed by more than 2 (out of 20) or when any attribute differed by 2 (out of 2). The attribute left versus right had the highest rate of disagreements by 2 (27% of all such disagreements), and substrate, overlap, and clarity together accounted for almost half of such disagreements (45%). Standardization of the definitions for each attribute and an increased familiarity with this rubric may reduce the variability among FFEs.

	Pattern	Quantity	Left vs Right	Contrast	Clarity	Distortion	Overlap	Substrate	Matrix	Suitability	Overall Score
FFE 1	1	0	0	1	0	0	0	1	0	1	**4**
FFE 2	0	0	0	2	0	0	0	0	0	0	**2**
Average	0.5	0	0	1.5	0	0	0	0.5	0	0.5	**3**

Table 2

Quality assessments for impression depicted in Figure 1.

	Pattern	Quantity	Left vs Right	Contrast	Clarity	Distortion	Overlap	Substrate	Matrix	Suitability	Overall Score
FFE 1	2	0	0	1	1	2	0	1	1	1	**9**
FFE 2	2	1	0	1	0	1	0	2	0	1	**8**
Average	2	0.5	0	1	0.5	1.5	0	1.5	0.5	1	**8.5**

Table 3

Quality assessments for impression depicted in Figure 2.

	Pattern	Quantity	Left vs Right	Contrast	Clarity	Distortion	Overlap	Substrate	Matrix	Suitability	Overall Score
FFE 1	1	2	2	1	1	2	1	1	1	1	**13**
FFE 2	2	2	1	2	1	1	1	1	1	1	**13**
Average	1.5	2	1.5	1.5	1	1.5	1	1	1	1	**13**

Table 4

Quality assessments for impression depicted in Figure 3.

	Pattern	Quantity	Left vs Right	Contrast	Clarity	Distortion	Overlap	Substrate	Matrix	Suitability	Overall Score
FFE 1	2	2	2	2	1	2	2	1	1	1	**16**
FFE 2	2	2	2	2	1	2	2	1	1	1	**16**
Average	2	2	2	2	1	2	2	1	1	1	**16**

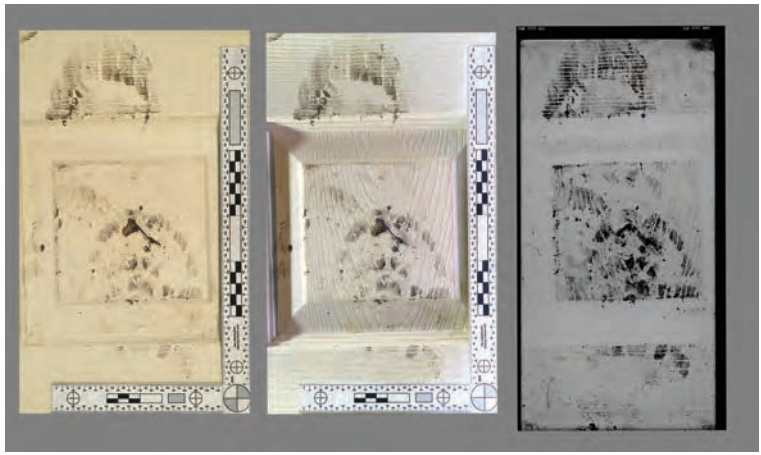
Table 5

Quality assessments for impression depicted in Figure 4.

	Pattern	Quantity	Left vs Right	Contrast	Clarity	Distortion	Overlap	Substrate	Matrix	Suitability	Overall Score
FFE 1	2	2	2	2	2	2	2	2	2	2	**20**
FFE 2	2	2	2	2	2	2	2	2	2	2	**20**
Average	2	2	2	2	2	2	2	2	2	2	**20**

Table 6

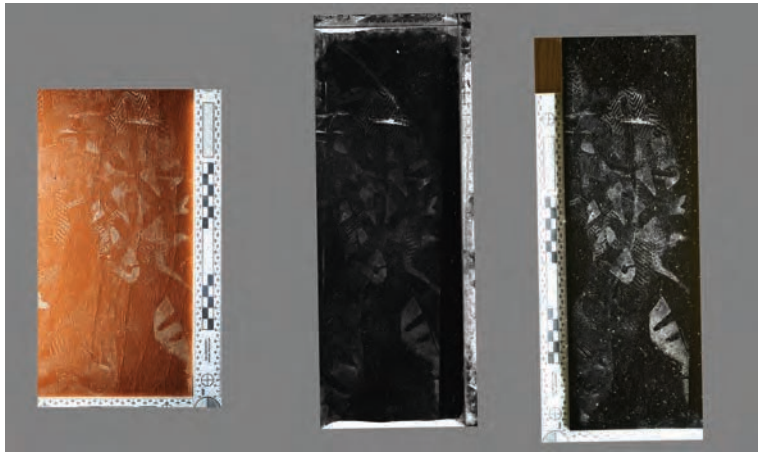
Quality assessments for impression depicted in Figure 5.



(a) *(b)* *(c)*

Figure 1

Soil impression on wooden door: (a) in-situ photograph, flood light; (b) in-situ photograph, oblique light; (c) gel lift scan.



(a) *(b)* *(c)*

Figure 2

Soil impression on hardwood flooring: (a) in-situ photograph, oblique light; (b) gel lift scan; (c) gel lift photograph, oblique light.

(a) *(b)*

Figure 3

Bloody impression on newspaper: (a) in-situ photograph, flood light; (b) in-situ photograph, flood light, after leucocrystal violet (LCV) application.

(a) *(b)*

Figure 4

Soil impression on ceramic tile: (a) in-situ photograph; (b) gel lift scan.

(a) *(b)*

Figure 5

Powdered residue impression on painted metal: (a) in-situ photograph, flood light; (b) gel lift scan.

Discussion and Conclusions

The method described here allowed us to characterize the quality of the questioned impressions generated for our FFE Black Box Study, which allowed us to control the distribution of a variety of attributes when selecting and assigning samples. We separated this as a standalone paper so that the FFE community can consider this as a potential basis for standardization. We recognize that if this is adopted informally by the community, or formally by a standards development organization, then a variety of revisions may of course be made as part of that process. We believe that with continued refinement via additional research and examiner feedback, this novel quality assessment tool holds the potential to mature into a standardized system for assessing the quality of questioned impressions. The attributes listed here may be useful in developing a common vocabulary among FFEs when assessing or describing footwear impressions or when assessing the suitability (value) of impressions for comparison. This framework may also hold the potential to inform the development of automated footwear impression quality assessment algorithms.

The forthcoming FFE Black Box Study report will further expound on the associations between these impression quality scores and the variability in participants' suitability determinations, conclusions, and reliability (i.e., intra-examiner repeatability and inter-examiner reproducibility).

For further information, please contact:

Brian C. McVicker
FBI Laboratory, Questioned Documents Unit
1205 Investigation Parkway
Quantico, VA 22135
bmcvicker@fbi.gov

Disclaimer

This is publication 20.80 of the FBI Laboratory Division. Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply endorsement of the manufacturer or its products or services by the FBI. The views expressed are those of the author(s) and do not necessarily reflect the official policy or position of the FBI, the U.S. Government, the Denver Police Department, or Noblis.

References

1. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council. *Strengthening Forensic Science in the United States*: A Path Forward; Document NO. 228091; National Academy Press: Washington, DC, 2009.
2. President's Council of Advisors on Science and Technology (PCAST). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*; Executive Office of the President, U.S. Government Printing Office: Washington, D.C., 2016.
3. Kellman, P. J.; Mnookin, J. L.; Erlikhman, G.; Garrigan, P.; Ghose, T.; Mettler, E.; Charlton, D.; Dror, I. E. Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates through Understanding and Predicting Difficulty. *PLOS One*. **2014**, *9* (5), e94617.
4. Ulery, B. T.; Hicklin, R. A.; Roberts, M. A.; Buscaglia, J. Factors Associated with Latent Fingerprint Exclusion Determinations. *For. Sci. Int*. **2017**, *275*, 65–75.
5. Hicklin, R. A.; Buscaglia, J.; Roberts, M. A. Assessing the Clarity of Friction Ridge Impressions. *For. Sci. Int*. **2013**, *266* (1–3), 106–117.

6. Scientific Working Group for Shoeprint and Tire Tread Evidence (SWGTREAD). Guide for the Examination of Footwear and Tire Impression Evidence. (3/2006) https://www.nist.gov/system/files/documents/2016/10/26/swgtread_08_guide_for_the_examination_of_footwear_and_tire_impression_evidence_200603.pdf Accessed June 1, 2021.
7. Su, H; Bouridane, A.; Crookes, D. Image Quality Measures for Hierarchical Decomposition of a Shoeprint Image. *For. Sci. Int*. **2006**, *163* (1–2), 125–131.
8. McElhone, R. L.; Meakin, G. E.; French, J. C.; Alexander, T.; Morgan, R. M. Simulating Forensic Casework Scenarios in Experimental Studies: The Generation of Footwear Marks in Blood. *For. Sci. Int*. **2016**, *264*, 34–40.
9. ISO (International Organization for Standardization) *Information Technology - Biometric Sample Quality - Part 1: Framework*; ISO/IEC 29794-1:2016; ISO: Geneva, Switzerland, 2016.