

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Examinations

Author(s): Igor Pacheco, Brian Cerchiai, Stephanie Stoiloff

Document No.: 248534

Date Received: December 2014

Award Number: 2010-DN-BX-K268

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

**Miami-Dade Research Study for the
Reliability of the ACE-V Process:
Accuracy & Precision in Latent Fingerprint Examinations**

Award Number 2010-DN-BX-K268

Final Technical Report

Submitted By:

Miami-Dade Police Department (MDPD) Forensic Services Bureau (FSB)

Igor Pacheco, CLPE (MDPD)
Brian Cerchiai, CTPE (MDPD)
Stephanie Stoiloff, M.S. (MDPD)

2014

Abstract

This research reports on an empirical study that evaluated the reliability of the Analysis, Comparison, and Evaluation (ACE) and Analysis, Comparison, Evaluation, and Verification (ACE-V) methodologies in latent fingerprint examinations. The participants’ performance was measured in terms of accuracy and precision, and was evaluated under both unbiased and biased conditions. Accuracy was measured in terms of the participant’s ability to correctly identify or exclude a latent print to a known source(s) and precision was measured in terms of the participant’s ability to reproduce and repeat the same conclusion. Reproducibility is defined as the ability of multiple participants to examine the same latent print and reach the same conclusion independently, while repeatability is defined as the participant’s ability to provide the same conclusion upon re-evaluation of the same latent print. For the purpose of this research, bias was defined as the ability of a participant to reproduce and repeat a conclusion when presented with two previous conclusions and asked to conduct a second verification.

The foundation of latent fingerprint identification is that friction ridge skin is unique and persistent. Through the examination of all of the qualitative and quantitative features available in friction ridge skin, impressions can be positively identified or excluded to the individual that produced it. This study reports the results of four categorical opinions: identification, exclusion, inconclusive, and no value decisions. In addition, sufficiency determinations and comparison decisions were evaluated based on a latent *Strength of Value* and *Difficulty of Comparison* rating scale that was designed for this research.

Tests were assembled using 80 latent prints with varying quantity and quality of information from ten known sources and were distributed to 109 latent print examiners across the United States. Participants had at least one year of latent print examination experience and employed the ACE methodology when comparing unknown latent prints to known sources. Responses from the participants yielded 5,963 sufficiency determinations, 4,536 ACE decisions, 532 ACE-V decisions, 1,311 repeatability decisions, 326 ACE decisions under biased conditions, and 333 repeatability decisions under biased conditions. This study took into account inconclusive responses in determining error rates and established a False Positive Rate (FPR) of 3.0% and False Negative Rate (FNR) of 7.5% for ACE examinations, as well as a FPR of 0.0% and FNR of 2.9% for ACE-V examinations. Participants were able to reproduce a correct identification 94.2% of the time and not reproduce an erroneous identification 100% of the time. Participants repeated their previous correct identifications 94.6% of the time and did not repeat their previous erroneous exclusions 93.1% of the time. Under biased conditions, participants were able to reproduce a correct identification 73.0% of the time and not reproduce an erroneous identification 96.5% of the time. Additionally, under biased conditions, participants repeated their previous correct identifications 93.2% of the time and did not repeat their previous erroneous exclusions 85.2% of the time.

This project was supported by Award No. 2010-DN-BX-K268 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the authors and do not necessarily reflect those of the Department of Justice.

Table of Contents

I. EXECUTIVE SUMMARY.....5

II. INTRODUCTION.....13

A. Purpose Goals and Objectives.....13

B. Background.....13

C. Review of Relevant Literature.....15

D. Research Questions and Hypotheses.....21

III. METHODS.....22

A. Experimental Design, Methods, and Materials.....22

B. Target Population.....34

C. Eligibility – Inclusion Criteria.....35

D. Accessible Population.....35

E. Sampling Plan & Setting.....35

F. Instrumentation.....36

G. Data Collection Methods.....36

H. Data Coding.....40

I. Descriptive Analysis.....44

J. Data Analysis Methods.....45

K. Definitions.....45

L. Internal Validity Strengths.....47

M. Internal Validity Weaknesses.....48

N. External Validity Strengths.....48

O. External Validity Weaknesses.....48

IV.	RESULTS.....	49
A.	Participant Background and Demographics.....	49
B.	Sufficiency Determinations.....	52
C.	Accuracy ACE & ACE-V.....	52
D.	Significantly Varied Results.....	55
E.	Precision (Reproducibility & Repeatability).....	58
F.	Bias (Reproducibility & Repeatability).....	61
V.	CONCLUSIONS.....	63
A.	Discussion of Findings.....	63
B.	Implications for Policy and Practice.....	69
C.	Implications for Further Research.....	70
VI.	REFERENCES.....	72
VII.	DISSEMINATION OF RESEARCH FINDINGS.....	73
VIII.	ACKNOWLEDGMENTS.....	73
IX.	APPENDIX A – Participant Exit Questionnaire.....	74
X.	APPENDIX B – Select Erroneous Identifications.....	75

I. EXECUTIVE SUMMARY

The purpose of this research was to conduct an empirical study to evaluate the reliability of latent fingerprint examiners using the Analysis, Comparison, and Evaluation (ACE) and Analysis, Comparison, Evaluation, and Verification (ACE-V) methodologies, as well as to determine various error rates of latent print examination decisions. The participants’ performance was measured in terms of accuracy and precision and evaluated under both unbiased and biased conditions.

The goal of this research was to determine if latent print examiners would be able to identify or exclude unknown latent prints to known sources using both the ACE and ACE-V methodologies. This research was conducted to answer the following research questions:

- Q1. Will participants be able to correctly identify or exclude unknown latent prints from known standards using the ACE methodology?
- Q2. Will participants be able to correctly identify or exclude unknown latent prints from known standards using the ACE-V methodology?
- Q3. Will participants reach significantly varied conclusions when comparing unknown latent prints to known standards using the ACE methodology?
- Q4. Will participants be able to reproduce and repeat conclusions from unknown latent prints to known standards using the ACE methodology?
- Q5. Will participants be able to reproduce and repeat conclusions from unknown latent prints to known standards using the ACE methodology under biased conditions?

This study utilized an experimental research design (Langenberg, 2009), and was conducted under testing conditions. Participants compared photographs of unknown latent prints to known standards, which consisted of ten sets of fingerprint and palm prints, to determine whether each of eighty unknown latent prints could be identified as having been made by a known source or excluded as having been made by all of the presented known sources. During Phase 1

and Phase 2 of this study, participants were instructed to make a comparison of unknown latent prints to three of ten standards that were provided and determine one of four categorical opinions: Identification, Exclusion, Inconclusive, or No Value. All participants in Phases 1 and 2 received identical tests. During Phase 3 of this study, participants were instructed to make a comparison of unknown latent prints to one of ten standards that were provided, and verify identifications, exclusions and inconclusives that were reported in Phases 1 and 2. Participants did not receive identical tests in this phase in order to test for bias and repeatability.

A panel of three International of Association (IAI) certified latent print examiners independently examined and compared 320 latent prints to known standards and scored each latent print and subsequent comparison and evaluation according to *Strength of Value* (source not present) and *Difficulty of Comparison* (source present) rating scales that were designed and used for this research. The three certified latent print examiners evaluated each latent based on three factors: a) minutiae b) minutiae formations and c) clarity (deposition pressure). Based on these factors, the *Strength of Value* and *Difficulty of Comparison* for each latent print was rated on a scale from 0-21 points and divided into three groups in order to present the participants with a broad range of latent print examinations that were representative of actual casework. These three groups were categorized as Insufficient to Difficult (0-7), Difficult to Moderate (>7-14), and Moderate to Easy (>14-21). The scores were independently determined by the three IAI certified latent print examiners for each latent print and comparison to its known standard; these scores were averaged, and 160 latent prints were selected based on similar difficulty ratings. Of these 160 latent prints, 80 were randomly selected for this study. In addition, 10 sets of known fingerprint and palm print standards were selected for testing purposes.

The participants in Phases 1, 2 and 3 of this study comprised a total of 76 different law enforcement agencies across the United States. The 109 participants represented 53 different local agencies, 16 different state agencies and 3 different federal agencies.

Participants reported a total of 5,963 sufficiency determinations. For latents with a *Strength of Value* rating in the 0-7 range, participants were in consensus and reported no value for identification 85.6% of the time, while latents in the >7-14 and >14-21 range were reported as of value for identification 88.1% and 99.6% of the time, respectively.

For participant performance related to ACE accuracy (Q1), the combined results of three categorical opinions (Identification, Exclusion, and Inconclusive) from Phase 1 and 2 ACE trials were evaluated. A total of 4,536 ACE examinations were reported by the participants, with a False Positive Rate (FPR) of 3.0% and a False Negative Rate (FNR) of 7.5% for ACE examinations. There were 42 erroneous identifications reported during ACE examinations. Although many of the errors appear to have been clerical in nature, the authors could not determine this with certainty.

For participant performance related to ACE-V accuracy (Q2), the combined verification results from Phase 3 ACE-V trials were evaluated. A total of 532 ACE-V examinations were reported by the participants, with a FPR of 0% and a FNR of 2.9% for ACE-V examinations. In comparing the number of reported erroneous identifications to erroneous exclusions for both ACE and ACE-V trials, the exclusion error rate was higher. In Phase 3, seventeen of the 42 erroneous identifications reported during ACE examinations were sent for verification to fourteen participants. None of the fourteen participants reported agreement with the initial erroneous identifications. During verification, the fourteen participants either reported that they disagreed or were inconclusive with the original conclusion. The remaining erroneous identifications reported

during ACE examinations were sent for verification under biased conditions and participants did reproduce or repeat at least one error.

To determine if the participants would reach significantly varied results using the ACE methodology (Q3), the combined results of three categorical opinions (Identification, Exclusion, and Inconclusive) from Phase 1 and 2 ACE trials were grouped by their *Difficulty of Comparison* (source present) and *Strength of Value* (source not present) ratings. For latents with a *Difficulty of Comparison* rating in the 0-7 range, participants were in consensus and reported an inconclusive result 76.4% of the time, while latents in the >7-14 and >14-21 range were reported as correct identifications 70.2% and 93.0% of the time, respectively. Additionally, for latents with a *Strength of Value* rating in the 0-7 range, participants were in consensus and reported an inconclusive result 68.2% of the time, while latents in the >7-14 and >14-21 range were reported as correct exclusions 64.8% and 83.6% of the time, respectively.

To determine if the participants would reproduce conclusions from comparisons of unknown latent prints to known standards made by other participants using the ACE methodology (Q4), the results of identification decisions from Phase 2 were sent to different participants in Phase 3 in order to determine if they would agree, disagree, or come to an inconclusive decision. The number of latent prints presented to participants for verification were based on 25 latent prints for a total of 532 participant verification decisions. The results indicate that the participants were able to reproduce a correct identification 94.2% of the time and not reproduce erroneous identifications 100% of the time.

To determine if the participants would repeat their own conclusions from comparisons of unknown latent prints to known standards using the ACE methodology (Q4), the results of identification decisions, erroneous exclusions, and inconclusive results where the source was

present from Phase 1 were sent to the same participants in Phase 3. The number of latent prints presented to participants for repeatability was based on 27 latent prints for a total of 1,311 participant decisions. The results indicate that participants repeated their previous correct identifications 94.6% of the time and did not repeat their previous erroneous identifications 68.8% of the time. Additionally, participants did not repeat their previous erroneous exclusions 93.1% of the time. The results from participants were almost evenly distributed when they were presented with their previous inconclusive decision and given the correct source, repeating their previous inconclusive decisions 49.4% of the time.

To determine if the participants would reproduce conclusions from comparisons of unknown latent prints to known standards made by other participants using the ACE methodology under biased conditions (Q5), the results of identification decisions from Phase 2 were sent to different participants in Phase 3 in order to determine if they would agree, disagree, or come to an inconclusive decision. The number of latent prints presented to participants for verification under biased conditions was based on 37 latent prints for a total of 329 participant verification decisions. Participants were able to reproduce a correct identification 73.0% of the time and not reproduce an erroneous identification 96.5% of the time.

To determine if the participants would repeat their previous conclusions from comparisons of unknown latent prints to known standards using the ACE methodology under biased conditions (Q5), the results of identification decisions, erroneous exclusions, and inconclusive results (where the source was present) from Phase 1 were sent to the same participants in Phase 3. The number of latent prints presented to participants for repeatability under biased conditions was based on 24 latent prints for a total of 333 participant decisions. Participants repeated their previous correct identifications 93.2% of the time (233 of 250 participant responses); repeated their previous

erroneous identifications 100% of the time (1 participant response) and did not repeat their previous erroneous exclusions 85.2% of the time. Additionally, 60% of the participants repeated their previous inconclusive decision when given the correct source.

A long standing issue within the latent fingerprint community is that latent print sufficiency determinations are not standardized in terms of a measurable scale. Due to the nature of friction ridge skin, transfer and collection of friction ridge skin impressions and human factors that exist during interpretation, sufficiency determinations should continue to be based on both quantitative and qualitative aspects. A *Strength of Value* rating scale, similar to the one designed for this research, could be utilized by latent print examiners in order to assist them in making appropriate sufficiency determinations.

In evaluating identification error rates within this study as it relates to participant demographics and the quantity and quality of information present during latent print comparisons, the data indicates an identification error rate decrease for participants with more latent print examination experience. However, identification error rates were nearly the same for participants with or without IAI latent print certification. Data also indicates that the identification error rate was less for the most difficult latent print comparisons, as none of the participants made an erroneous identification and reported more inconclusive decisions.

In evaluating exclusion error rates within this study as it relates to participant demographics and the quantity and quality of information present during latent print comparisons, the data indicates that exclusion error rates were higher than identification error rates irrespective of the participant's years of latent print examination experience. Additionally, exclusion error rates did not change for participants with or without IAI latent print certification. For exclusions error rates

related to the difficulty of the latent print comparison, the exclusion error rate was lowest for latent trials that were rated the easiest to compare.

Although this study was not designed to precisely measure how the participants applied the ACE methodology when making their comparisons, data was collected from the participant answer sheets that indicated key components of their Analysis (clarity, anatomical source and certainty of orientation), Comparison (standards used in their comparison), and Evaluation (identification, exclusion, or inconclusive). When taking into account the error rate of erroneous identifications and erroneous exclusions during ACE trials, the findings support the importance of an independent review of fingerprint conclusions to reduce errors in fingerprint examinations.

Different trials were sent to participants for verification and second verification reproducibility trials, and the overall results indicate that a contextual bias may have been introduced when participants were presented with two previous conclusions and asked to perform a second verification. Participants who were asked to perform a second verification agreed less often with an initial correct identification and reported more inconclusive decisions. In addition, when participants performing a second verification were presented with an initial erroneous identification, participants reported less inconclusive decisions and were more likely to either agree or disagree with an incorrect identification.

Different trials were also sent to participants to test for repeatability in the form of a verification and second verification. The participants were not made aware that they were verifying or conducting a second verification of their previous conclusions. In testing for repeatability, the effects of contextual bias may have also been introduced when participants were presented with two previous conclusions and asked to perform a second verification. Participants performing second verifications repeated their previous erroneous exclusions and inconclusive decisions more

often, and were less likely to change these decisions to correct identifications. However, when participants were presented with their initial correct identifications as a second verification, contextual bias did not appear to be a factor as all participants repeated their initial correct identifications at approximately the same rate.

All of the data from this study could not be captured for the purposes of this report. The authors have chosen to report the most significant findings as stated in the initial research proposal.

This research project was sponsored by the National Institute of Justice of the Department of Justice of the United States of America under Award No. 2010-DN-BX-K268.

II. INTRODUCTION

A. Purpose, Goals and Objectives

The purpose of this research was to conduct an empirical study and evaluate the reliability of latent fingerprint examiners using the Analysis, Comparison, Evaluation, and Verification (ACE & ACE-V) methodology, as well as to determine various error rates of latent fingerprint examination identification and exclusion decisions. Participant performance was measured in terms of accuracy and precision and evaluated under both unbiased and biased conditions.

The goal of this research was to determine if latent print examiners could be able to identify or exclude unknown latent prints to known sources using both the ACE and ACE-V methodologies under various conditions.

B. Background

Fingerprint conclusions have traditionally been based on an arbitrary number of characteristics, commonly called points, to determine a match. In 1914, fingerprint pioneer and French anthropologist Dr. Edmond Locard published his conclusions stating that a minimum of eight points was required for positive fingerprint identifications (Ashbaugh, 1999). Since 1973, the International Association for Identification (IAI) has stated that, “no valid basis exists at this time for requiring that a pre-determined minimum number of friction ridge characteristics must be present in two impressions in order to establish positive identification.” Furthermore, the Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) has stated that, “friction ridge impression examinations are conducted by examiners using the Analysis, Comparison, Evaluation, and Verification (ACE-V) methodology, which include both qualitative and quantitative aspects.”

The ACE methodology was first articulated by R.A. Huber of the Royal Canadian Mountain Police (RCMP) in 1959. ACE is not limited to fingerprint examinations, but can be applied in general when comparing two or more objects. In 1972 Huber stated, “The process has 3 distinct stages through which one must pass consciously or unconsciously in the course of an examination. The Analysis stage is when the unknown item must be reduced to a matter of its properties, or characteristics. These properties may be directly observable, measurable, or otherwise perceptual qualities. The Comparison stage is when the properties or characteristics of the unknown, determined through analysis, are now compared with the familiar or recorded properties of known items. The Evaluation stage is when similarities or dissimilarities in properties or characteristics will each have a certain value for identification purposes determined by its likelihood of occurrence. The weight or significance of each must therefore be considered.” For purposes of peer review in latent print examinations, a fourth step, the Verification stage, has been added and involves an independent ACE process by another examiner. The ACE process including the verification stage is what is commonly referred to as the ACE-V method.

The scientific approach of the ACE-V process was detailed in an article by the Federal Bureau of Investigation and key points are summarized below:

- 1) Make an initial observation

An examiner observes friction ridge detail on an item of evidence.

- 2) State the problem or question

Who is the source of this latent print?

- 3) Generate a hypothesis

H₀ *The latent print did not originate from the same source as the known print.*

H₁ *The latent print did originate from the same source as the known print.*

4) Conduct tests

The Analysis phase involves gathering all of the information available in both the known and unknown prints to determine if sufficient quality and quantity exist for the print to be individualized. If the examiner determines that the information in the unknown print is of sufficient quality and quantity, the print is declared “of value” for individualization. The Comparison phase entails examining the information gathered in the analyses of the two prints to discern similarities and differences in their friction ridge arrangements.

5) Generate conclusions based on the data

After fully comparing the two prints, the examiner can reach conclusions based on all of the information present. This is the Evaluation phase of ACE-V.

6) Confirm the process and conclusion through repetition (replication) by others

In the Verification phase, another examiner performs an independent analysis, comparison, and evaluation of the two prints in question. This is akin to replication of an experiment to verify the results that were obtained.

7) Record and/or present the conclusions

Examiners report and/or present the results of their examinations through written communications or oral testimony.

(Forensic Science Communications, October 2009, Vol. 11, No. 4)

C. Review of Relevant Literature

A review of the relevant literature provides recent studies that address the reliability of fingerprint evidence. These studies evaluate the accuracy and precision of conclusions resulting from fingerprint examination comparisons of unknown latent prints to known standards under unbiased and biased conditions.

Evetts and Williams (1995) published the first study in which they reviewed fingerprint conclusions based on a sixteen point standard used in the United Kingdom. A total of 130 fingerprint experts from bureaus in England and Wales with ten or more years of experience conducted independent comparisons of ten latent impressions to known standards. Nine of these latent impressions were from provided known sources and one latent impression was from a source

not present. Each expert was asked to decide whether the result was: a) a full identification b) a non-provable identification c) not-identical or d) insufficient detail for an opinion. Overall, no erroneous identifications and ten erroneous exclusions were reported by the participants.

Wertheim, Langenburg, and Moenssens (2006) published data from a training environment. A total of 108 participants performed comparisons from training packets containing ten latent impressions and eight known standards. The training courses were opened to participants of any skill level, including participants with no training and experience. As a result, the authors separated the data of participants with more than one year of experience from the data of participants with one year of experience or less. All of the latent impressions in the study had the source present in the standards. Of the 108 participants, 92 reported more than one year of experience and sixteen of the participants reported they either had no training and experience or less than a year of experience. The 92 examiners with more than one year of experience made a total of 5,861 identifications in which 61 erroneous identifications were recorded. Two of the decisions were believed to be true erroneous identifications and the other 59 were determined to be clerical errors. This resulted in an ACE erroneous individualization rate of 0.034% and a clerical error rate of 1.01% for the participants with more than one year of experience during these training exercises. A follow up verification study was also performed of the errors reported by previous participants. Sixteen participants with more than one year of experience acted as verifiers to previous participants' results. Each verifier was given a packet to verify containing the results of eight correct individualizations and two errors. The sixteen independent verifiers did not verify any of the errors given to them in the verification packet exercises. Additionally, the sixteen participants with no training and experience or less than a year of experience in the original study did commit significantly more errors than the experienced group.

Langenburg, Champod, and Wertheim (2009) reported data from a series of tests under various levels of contextual bias. Their experiment was partly conducted at the International Association for Identification 91st Educational Conference in Boston, Massachusetts. The authors solicited fingerprint experts attending the conference and a total of 43 experienced examiners participated (1-29 years). Unbeknownst to the participants, the authors separated them into three different groups: a control group, low bias group and a high bias group. Each group was given a set of six side-by-side comparisons (Q1-Q6) of a latent print and known exemplar. Participants were instructed to provide opinions of individualization, exclusion, or inconclusive. Additionally, the participants were asked to count the number of minutiae in agreement and disagreement and to rate the clarity of the both latent prints and known exemplars. The control group received comparisons with no contextual information; the low bias group received comparisons in which a conclusion was already provided and asked to render a decision of whether they agreed or disagreed with the conclusion; and the high bias group was presented with the same information/instructions as the low bias group, however, this group was also told by a prominent internationally recognized fingerprint expert that these were his opinions from an actual case. Subsequently, the authors repeated the same experiment with 86 lay persons who had no training or experience in conducting fingerprint examinations. The lay persons were University students who attended a community college in St. Paul, Minnesota and ranged in age from 19 to 65 years old. For the six trials (Q1-Q6) that were presented to both the expert and novice participants, three trials were classified as being from the same source and three trials were classified as being from a different source. The six trials ranged in comparison difficulty (easy, medium, to difficult) and one trial (Q2) was considered to be a close non-match as a result of a previous Automated Fingerprint Identification System (AFIS) database search. For both expert and novice low and high

bias groups, the participants were prompted with four trials that were presented to them as individualizations, one trial as an exclusion, and one trial as an inconclusive. The results showed that fingerprint experts were influenced by contextual information (previous conclusion provided) during fingerprint comparisons, but not towards making errors (individualizations and exclusion decisions). Instead, fingerprint experts under biasing conditions provided significantly fewer definitive and erroneous conclusions than the control group. In contrast, the novice participants were more influenced by the biased conditions and did tend to make incorrect decisions, especially when prompted towards an incorrect response by the biasing condition. Additionally, the fingerprint experts committed far fewer errors (three erroneous exclusions and one erroneous identification) than did the novices. The four errors that were committed by the experts all occurred in the control group with no biasing condition and the one erroneous identification was not from the trial that was a close non-match (Q2). In contrast, the novice participants committed 24 erroneous individualizations (seven in the control group, seven in the low bias group, and ten in the high bias group). Of the ten erroneous individualizations in the high bias group, nine of the errors were committed in trial Q2. Moreover, the novice participants also committed 22 erroneous exclusions (eleven in the control group, ten in the low bias group, and one in the high bias group).

Langenburg (2009) published data from a method performance study of the ACE-V process. The study tested the accuracy, precision, reproducibility, repeatability, and biasability that result from the ACE-V process. A total of six experienced examiners from the same agency laboratory (ranging from 6 – 35 years) performed comparisons of 120 latent impressions to eight known standards. The study was separated into three phases and resulted in 60 ACE and 60 ACE-V trials per participant. Of these trials, no erroneous identifications and 32 erroneous exclusions were reported. The results showed a high degree of accuracy with respect to opinions where

identification was reported, but lower accuracy with respect to opinions of exclusion. Overall, the participants in this study were generally consistent with respect to the number of categorical opinions for the ACE and ACE-V trials.

Ulery, Hicklin, Buscaglia and Roberts (2011) published a large scale study on the accuracy and reliability of forensic latent fingerprint decisions. A total of 169 latent print examiners each compared approximately 100 pairs of latent and exemplar prints from a pool of 744 pairs, which were assigned at random. The participants varied with respect to their organization, training history, and other demographics; in general, the group of participants was highly experienced. The median number of years of experience was ten, and 83% of the participants were certified latent print examiners. The 744 latent-exemplar pairs included 356 latents from 165 distinct fingers from 21 people, and 484 exemplars. A total of 520 mated and 224 nonmated pairs, were utilized. The participants made a total of 17,121 decisions and 23% of all decisions resulted in “no value” decisions. A total of six erroneous identifications (0.1%) occurred among 4,083 latents that were deemed of value for identification. The six errors were committed by five examiners, three of whom were certified, one who was not certified, and the other was not known (one certified examiner made two erroneous identifications). In addition, participants reported 450 erroneous exclusions (7.5%) among 5,969 latents that were deemed to be of value for identification. At least one erroneous exclusion was reported by 85% of the participants.

In a follow-up to their initial study, Ulery, Hicklin, Buscaglia and Roberts (2012) reported results on the repeatability and reproducibility of decisions by latent print examiners. Of the 169 examiners who participated in their initial study, 72 examiners were presented with the same prints after a seven month interval to determine if one examiner would consistently reach the same decision on the same fingerprints (repeatability), although they were not told they had previously

seen these prints. Data was also reported from the same 72 examiners’ initial test results to determine whether different examiners had reached the same decision on the same fingerprints (reproducibility). Each examiner in the repeatability retest was assigned 25 comparisons from a pool of 744 image pairs. Latent print examiners repeated 89.1% of their individualizations and 90.1% repeated their exclusions; no false positive errors were repeated, and 30% of false negative errors were repeated. Most of the changed decisions resulted in inconclusive decisions. Repeatability of all comparison decisions combined was 90.0% for mated pairs; 85.9% for non-mated pairs.

To compare repeatability results to the results for reproducibility, responses from the 72 examiners’ initial tests were taken for approximately 100 image pairs. The authors used a percentage agreement, \overline{P} , to describe both intra-examiner agreement (repeatability) and inter-examiner agreement (reproducibility). The intra- \overline{P} for comparison decisions of individualization, any exclusion, and other was 90.3% and 85.9% for mated and non-mated pairs respectively. In contrast, the inter- \overline{P} for comparison decisions of individualization, any exclusion, and other was 79.8% for mated pairs and 79.6% for non-mated pairs. In addition, decreased repeatability and reproducibility for both individualizations and exclusion decisions appeared to be related to the difficulty of the comparison.

The research suggests that fingerprint examinations by trained examiners result in few false positives and false negative conclusions. However, more studies are needed to further investigate the reliability of all conclusions by examiners applying the ACE process, including under different types of bias. In addition, studies that include a large sample size across multiple laboratories/agencies nationwide would be more representative of the fingerprint community versus results that are obtained from a single laboratory/agency.

D. Research Questions and Hypothesis

The proposed outcome in this section is presented with the intention that the findings will be able to answer the following research questions:

Research Questions

Q1. Will participants be able to correctly identify or exclude unknown latent prints from known standards using the ACE methodology?

Q2. Will participants be able to correctly identify or exclude unknown latent prints from known standards using the ACE-V methodology?

Q3. Will participants reach significantly varied conclusions when comparing unknown latent prints to known standards using the ACE methodology?

Q4. Will participants be able to reproduce and repeat conclusions from unknown latent prints to known standards using the ACE methodology?

Q5. Will participants be able to reproduce and repeat conclusions from unknown latent prints to known standards using the ACE methodology under biased conditions?

Research Hypotheses

H1. Participants will be able to correctly identify or exclude unknown latent prints from known standards using the ACE methodology.

H2. Participants will be able to correctly identify or exclude unknown latent prints from known standards using the ACE-V methodology.

H3. Participants will not reach significantly varied conclusions when comparing unknown latent prints to known standards using the ACE methodology.

H4. Participants will be able to reproduce and repeat conclusions from unknown latent prints to known standards using the ACE methodology.

H5. Participants will be able to reproduce and repeat conclusions from unknown latent prints to known standards using the ACE methodology under biased conditions.

There were several dependent variables that were examined in this study. The first dependent variable was the accuracy of correct identifications and correct exclusions, which was

measured by whether or not the unknown latent print could be correctly identified or correctly excluded to the known standard by comparing the quality and quantity of information present in both impressions. The second dependent variable was the precision of categorical opinions, which was measured by the reproducibility and repeatability of participants’ conclusions.

There are several independent variables in this study, such as the experience and qualifications of the examiner, analysis of the latent print, difficulty of the latent print comparison, and the effects of bias. For Q1, Q2 and H1, H2 the researchers were interested in evaluating the overall accuracy of ACE and ACE-V. For Q3, H3 the researchers were interested in determining if the participants would reach significantly varied results. For Q4, H4 the researchers were interested in evaluating the reproducibility and repeatability of categorical opinions. For Q5, H5 the researchers were interested in evaluating the reproducibility and repeatability of categorical opinions under biased conditions.

III. METHODS

A. Experimental Design, Methods and Materials

This study utilized an experimental research design (Langenberg, 2009), and was conducted under testing conditions. Participants compared photographs of unknown latent prints to ten sets of known fingerprint and palm print standards, to determine whether each of eighty unknown latent prints could be identified as having been made by a known source or excluded as having been made by all of the presented known sources. During Phase 1 and Phase 2 of this study, participants were instructed to make a comparison of unknown latent prints to three of ten standards that were provided and determine one of four categorical opinions: Identification, Exclusion, Inconclusive, or No Value. During Phase 3 of this study, participants were instructed

to make a comparison of unknown latent prints to one of ten standards that were provided, and verify identifications, exclusions and inconclusives that were reported in Phases 1 and 2 under both unbiased and biased conditions.

Fingerprint and palm print standards and latent impressions were created from thirteen volunteer sources. All of the volunteer sources signed a consent form to participate and these forms and associated identifying information were kept confidential. The volunteers consisted of men and women who were chosen primarily due to not having their fingerprints and palm prints in any known AFIS database outside of Miami-Dade County (MDC). This selection method protected the integrity of our research results, as participants could not utilize an AFIS search to assist them in identifying any of the latent prints provided to them within this study.

In creating the test sets for this study, one (1) fingerprint and two (2) palm print standards were collected from each volunteer using an AFIS MorphoTrak LSS-3000R livescan and printed on Federal Bureau of Investigation (FBI) FD-249 fingerprint cards and FD-884 palm print cards from a Lexmark T522 printer. A combined total of 2,711 latent photographs were also created from latent prints deposited by each volunteer and collected from non-porous, flat and curved surfaces (plastic, tile, metal, and glass). Latent prints were lifted using regular black powder, tape, and white backing cards. The latent lifts were then scanned into Adobe Photoshop at 1000ppi using an Epson Perfection 4990 Photo Scanner. The latent prints were printed at a 1:1 scale on Kodak Royal Digital Photo Paper (F surface) from a Noritsu QSS-3212 printer. Each latent photograph contained a single latent print. To ensure anonymity of the volunteers and track the ground truth (known source) of each latent impression, each volunteer was assigned a pseudonym from the Greek alphabet. The number of latent prints provided by each volunteer was as follows: Alpha-

130; Beta-186; Delta-346; Epsilon-280; Eta-161; Gamma-257; Iota-245; Kappa-241; Lambda-105; Mu-256; Nu-105; Theta-107; Zeta-292.

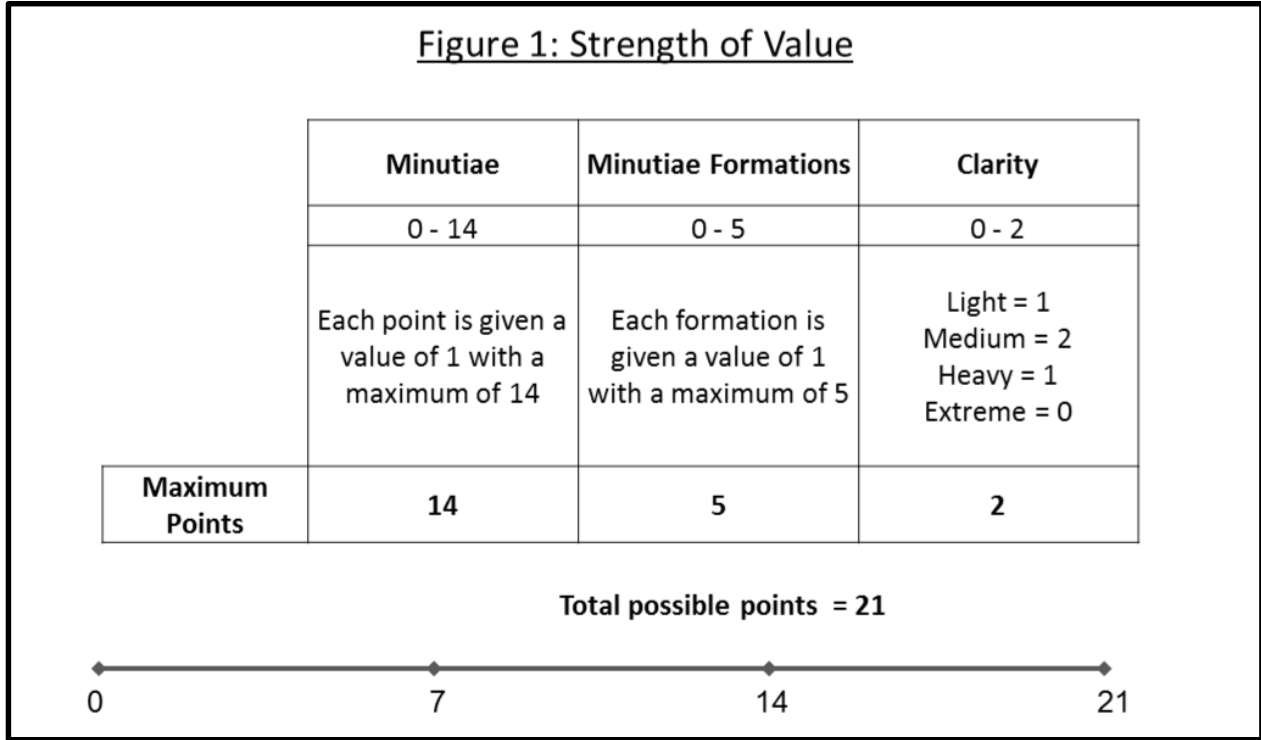
From the 2,711 latent prints that were created, 320 latent prints with varying quality and quantity of information were chosen by the researchers. A panel of three International of Association (IAI) certified latent print examiners independently examined and compared the 320 latent prints to the known standards and scored each latent print and subsequent comparison to their known standard according to a rating scale that was designed and used for this research; 80 were selected as the final latent prints to be used for testing purposes.

The three certified latent print examiners rated the following factors:

- Strength of Value of Latent Print
- Latent in Agreement with Standard
- Difficulty of Comparison

Strength of Value of Latent Print

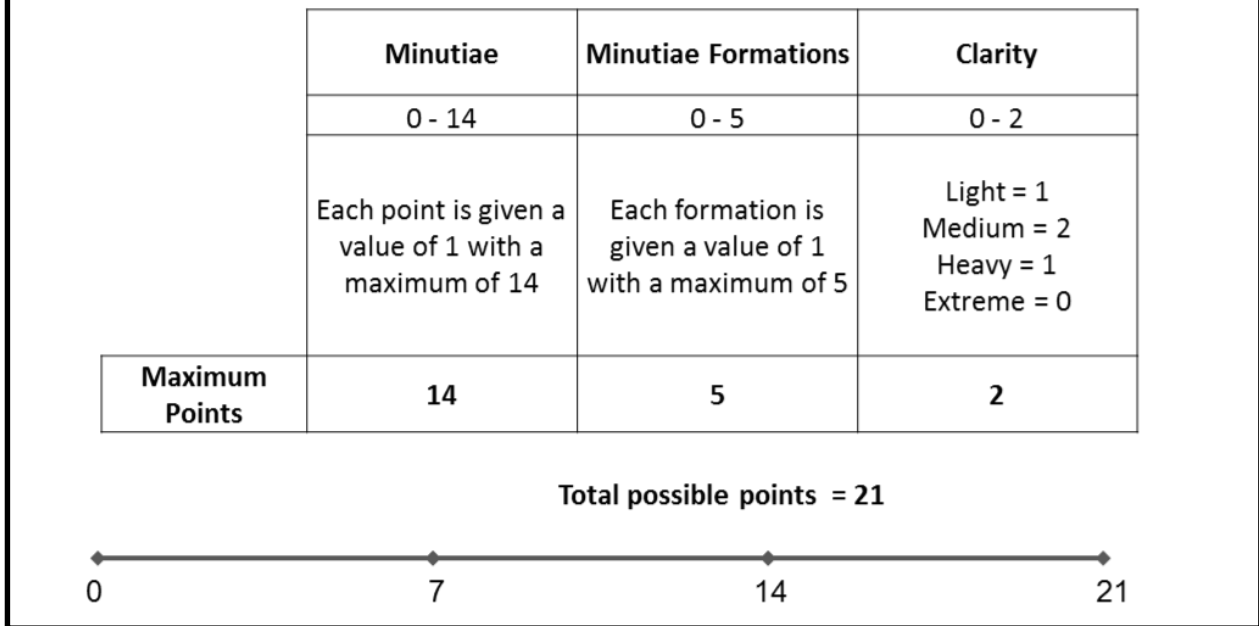
To determine the *Strength of Value* of the latent print rating, the panel of three IAI certified latent print examiners evaluated each latent based on three factors: a) minutiae b) minutiae formations and c) clarity (deposition pressure). Based on these three factors, the *Strength of Value* for each latent was rated on a scale from 0-21 points (Figure 1). For minutiae present, one point was given for each visible minutiae characteristic (i.e., bifurcations, ending ridges, and dots) with a maximum of 14 points. For minutiae formations present, one point was given for each visible minutiae formation (i.e., enclosures and rows of dots) with a maximum of 5 points. For clarity, each latent was scored based on the deposition pressure of the impression: Light – 1, Medium – 2, Heavy – 1, and Extreme – 0, for a maximum of 2 points.



Latent in Agreement with Standard

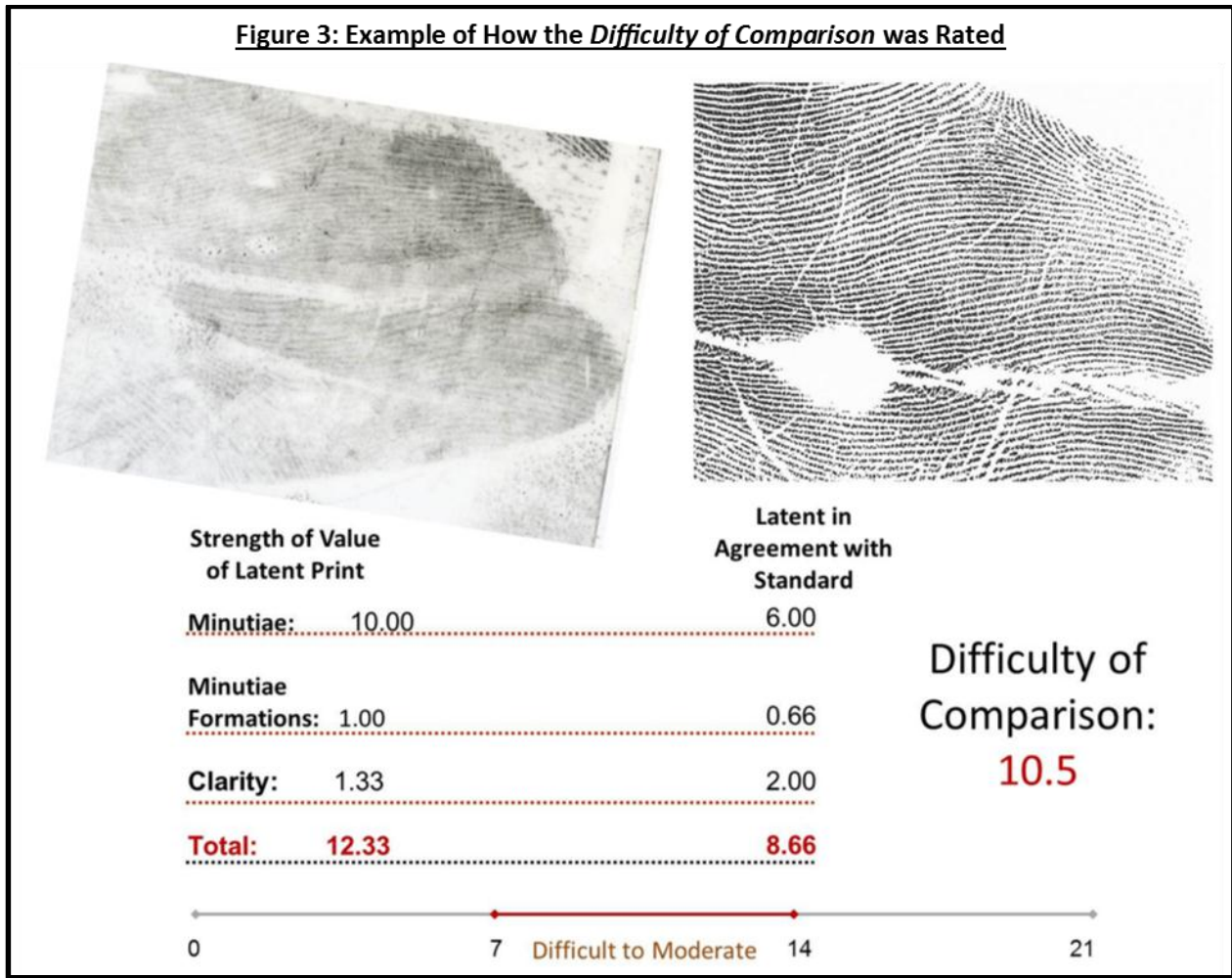
To determine the *Latent in Agreement with Standard* rating, the panel of three IAI certified latent print examiners compared each latent print to its known standard to determine if the minutiae and minutiae formations were in agreement (present in both). Clarity was also assessed in the known standard and all three factors were rated in the same manner as in the *Strength of Value* of the latent print rating on a scale from 0–21 points (Figure 2).

Figure 2: Latent In Agreement with Standard



Difficulty of Comparison

To determine the *Difficulty of Comparison* rating, an average score was calculated from both the *Strength of Value* of the latent print and the *Latent in Agreement with Standard* ratings. The lower the score, the more difficult the comparison was rated. The *Difficulty of Comparison* scale ranged from 0-21 points. An example is shown in Figure 3.



The scores were independently determined by the three IAI certified latent print examiners for each latent print and comparison to its known standard. These scores were averaged and 160 latent prints were selected based on similar difficulty ratings. The 160 latent prints were then divided into the following groups in order to present the participants with a broad range of latent print examinations that were representative of actual casework:

- **112 Latent Prints with Source Present (70%)**
 - 28 Insufficient to Difficult (25%)
 - 42 Difficult to Moderate (37.5%)
 - 42 Moderate to Easy (37.5%)

- **48 Latent Prints with Source Not Present (30%)**
 - 12 Insufficient to Difficult (25%)
 - 18 Difficult to Moderate (37.5%)
 - 18 Moderate to Easy (37.5%)

Of the 112 latent prints with the source present, 56 were randomly selected for this study. Similarly, for the 48 latent prints with the source not present, 24 were randomly selected for this study. In total, 80 latent prints were used for this research. In addition, of the thirteen known volunteer sources, only ten sets of fingerprint and palm print standards were selected for testing purposes. The final distribution of latent prints and known fingerprint and palm print standards are described in Table 1.

Table 1: Final Distribution of Latent Prints and Sources with Difficulty Range

		Known Sources	Insufficient to Difficult	Difficult to Moderate	Moderate to Easy
Source Present	Alpha		4	4	1
	Beta		2	4	4
	Gamma		2	5	2
	Delta		1	0	3
	Epsilon		0	1	1
	Zeta		1	3	3
	Eta		1	1	0
	Theta		1	1	1
	Iota		1	1	4
	Kappa		1	1	2
Source Not Present	Lambda*		1	2	4
	Mu*		4	2	1
	Nu*		1	5	4
Totals:			20	30	30
			25%	37.5%	37.5%
			25%	75.0%	

Source Present	70.00%	56
Insufficient to Difficult	14	25%
Difficult to Moderate	21	37.5%
Moderate to Easy	21	37.5%

Source Not Present	30.00%	24
Insufficient to Difficult	6	25%
Difficult to Moderate	9	37.5%
Moderate to Easy	9	37.5%

Total Latents:
80

**Note: The Strength of Value of Latent Print rating was used for latent prints that came from known sources that were not presented to the participants for comparison.*

Phase 1

In Phase 1 of this research study, each participant was mailed ten sets of known fingerprint and palm print standards, 40 latent prints, an answer sheet, and a return envelope. All participants received identical tests. The participants were instructed to perform an ACE examination for each latent print and make a comparison to three of the ten sets of standards provided.

Analysis - Participants were to perform an analysis of each latent print and asked to indicate the clarity, anatomical source, and orientation, as well as whether they determined the latent print to be of value or no value for identification. Clarity was to be indicated by the level of friction ridge detail present in the latent print:

- Level 1: Overall ridge flow
- Level 2: Individual friction ridge paths, friction ridge events (e.g., bifurcations, ending ridges, dots, and continuous ridges) and their relative arrangements
- Level 3: Ridge structures (edge shapes, pores) and their relative arrangements

Creases, scars, warts, incipient ridges, and other features may have been reflected in all three levels of detail. The anatomical source was to be indicated by the participant’s determination of the source of the latent print as a Fingerprint (FP) or Palm Print (PP). If the anatomical source could not be determined, the print was to be marked as an Impression (IMP). Orientation was indicated by the participant’s certainty or uncertainty of the proper direction of an area of friction ridge detail. After evaluating these three factors, the participants were asked to indicate if the latent print was of value or no value and instructed that a value determination was only to be marked when they believed that the quantity and quality of information was sufficient in the latent print in that an identification could be made. Furthermore, for the purposes of this study, the participants

were told that latent prints insufficient for identification, but sufficient for exclusion, should be marked as no value.

Comparison - After conducting their analysis, the participants were instructed to compare each latent print to three sets of fingerprint and palm print standards that were listed for each latent trial.

Evaluation - After conducting a comparison of the latent prints to the three sets of standards, the participants were asked to indicate whether they had made an identification, an exclusion, or reached inconclusive decision. An identification was to be reported if the participant determined that there was sufficient quality and quantity of detail in agreement to conclude that two areas of friction ridge impressions did originate from the same source. Identifications were documented by indicating the standard, finger number, right palm or left palm. An exclusion was to be reported if the participant determined that there was sufficient quality and quantity of detail in disagreement to conclude that two areas of friction ridge impressions did not originate from the same source. Lastly, an inconclusive result was to be reported if the participant determined that there was neither sufficient agreement to individualize, nor sufficient disagreement to exclude the latent print to the three known standards. If the participant indicated they reached an inconclusive result, they were asked to indicate why by marking one of the following reasons:

- Poor quality standards
- Sufficient detail for comparison, but insufficient to identify
- Cannot fully exclude all three individuals
- Other: Brief explanation

Documentation of the ACE examination was captured as shown in Figure 4.

Figure 4: ACE Trial

Sample	ANALYSIS				COMPARISON	EVALUATION													
	Clarity	Anatomical Source:	Orientation:	Value		#1	#2	#3	#4	#5	Right Palm	#6	#7	#8	#9	#10	Left Palm	Exclusion	Inconclusive
1	<input type="checkbox"/>	FP <input type="checkbox"/>	Certain <input type="checkbox"/>	<input type="checkbox"/>	Compare to 3 Standards A, B, C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2	<input type="checkbox"/>	PP <input type="checkbox"/>	Uncertain <input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	IMP <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<i>If Inconclusive, indicate reason:</i> <input type="checkbox"/> Poor Quality Standards <input type="checkbox"/> Sufficient detail for comparison, but insufficient to identify <input type="checkbox"/> Cannot fully exclude all three individuals <input type="checkbox"/> Other:													

Phase 2

In Phase 2 of this research study, the participants were randomly divided among two groups, Group A and Group B. The remaining 40 latent prints that were not used in Phase 1 were divided into two groups of 20. Thus, Group A participants were mailed ten sets of known fingerprint and palm print standards, 20 latent prints, an answer sheet and a return envelope; and Group B participants were mailed ten sets of fingerprint and palm print standards, 20 latent prints, an answer sheet, and a return envelope. As in Phase 1, the participants in Group A and Group B were given identical tests, respectively, and instructed to perform an ACE examination for each latent print and make a comparison to three of the ten sets of standards provided. The answer sheet and instructions for Phase 2 were the same as were presented to the participants in Phase 1.

Phase 3

In Phase 3 of this research study, participants were divided into 2 subgroups. Both subgroups were mailed ten sets of known fingerprint and palm print standards, a set of select latent prints, an answer sheet, and a return envelope. Subgroup 1 participants were instructed to perform a verification for each latent print and make a comparison to one of the ten sets of standards provided. In order to test for bias and repeatability, the participants did not receive identical tests.

For an *Identification*, participants were presented with an identification to a specific standard and a specific area (finger/palm).

For a *Verification*, participants were asked to verify the Identification and instructed to mark the “Agree” or “Disagree” box. If the participant determined that there was neither sufficient agreement to individualize, nor sufficient disagreement to exclude, they were instructed to mark the “Inconclusive” box. Any relevant information supporting their conclusion was to be recorded in the “Comments” section.

Each participant in Subgroup 1 was sent a test that included latent verification trials where the participant was presented with another participant’s correct and erroneous identifications from Phase 2. In addition, tests included latent verification trials where the participant was re-presented with at least one of their own trials with a previous conclusion: correct identification, erroneous identification, erroneous exclusions, or inconclusive results (where the source was present) from Phase 1 in order to test for repeatability. The participants were not made aware that they were verifying their own answers. An example of the documentation of a verification trial is represented in Figure 5.

Figure 5: Verification Trial

Sample Latent # 1	Identification	Second Verification		
	Has been identified to:	Agree <input type="checkbox"/>	Disagree <input type="checkbox"/>	Inconclusive <input type="checkbox"/>
	Standard A Finger #6	Comments:		

For Subgroup 2, participants were instructed to perform a *second* verification for each latent print and to make a comparison to one of the ten sets of standards provided. In order to test for bias and repeatability, the participants did not receive identical tests.

For an *Identification*, participants were presented with an identification to a specific standard and a specific area (finger/palm).

For a *Verification*, participants were presented with a verification to the same specific standard and specific area (finger/palm) as the identification.

For a *Second Verification*, participants were asked to perform a *second* verification of the *initial* Identification and instructed to mark the “Agree” or “Disagree” box. If the participant determined that there was neither sufficient agreement to individualize, nor sufficient disagreement to exclude, they were instructed to mark the “Inconclusive” box. Any relevant information supporting their conclusion was to be recorded in the Comments section. An example of the documentation of a Second Verification trial is represented in Figure 6.

Figure 6: Second Verification Trial

	Identification	Verification	Second Verification
Sample Latent # 2	Has been identified to:	Has been verified to:	Agree Disagree Inconclusive <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
	Standard	Standard	
	C Finger #4	C Finger #4	Comments:

Each participant in Subgroup 2 was sent a test that included latent verification trials where the participant was presented with another participant’s correct identifications and erroneous identifications from Phase 2 and a verification conclusion. In order to test for bias, the participant

was asked to perform a *second* verification. In addition, their tests included latent verification trials where the participant was re-presented with at least one of their own trials with a previous conclusion from Phase 1 (correct identification, erroneous identification, erroneous exclusion, or inconclusive results where the source was present) in order to test for repeatability under biased conditions. The participants were not made aware that they were verifying their own answers.

B. Target Population

In this study, the target population represented a subset of the forensic science community; more specifically, the target population were individuals who were currently employed by a law enforcement agency (crime laboratory), or like agency, in the United States and performing latent print examinations. Retired or contracted latent print examiners were also eligible to participate. The Miami-Dade Police Department (MDPD) Forensic Services Bureau (FSB) utilized the Active Membership list of latent print examiners in the IAI.

In addition, the Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) establishes guidelines for the qualifications and training to competency of latent print examiner trainees. These guidelines state that all latent print examiner trainees should receive a minimum of one year of full-time latent print work experience, with the majority of the time spent on the analysis, comparisons, and evaluation of impressions. Additionally, SWGFAST recommends that latent print examiner trainees receive two or more years of full-time latent print work with the majority of the time spent on the analysis, comparison, and evaluation of impressions in order to demonstrate competency in friction ridge examination. Latent print examiners in the United States who were an active member of the IAI received an email invitation from the MDPD FSB inviting them to participate in this study; this invitation included the

completion of a short questionnaire including qualifications and experience. The test sets utilized in this study were similar to the work that participants perform on a daily basis.

C. Eligibility – Inclusion Criteria

Eligible participants were required to have one (1) year of active casework experience using the ACE-V methodology as a latent print examiner at a law enforcement agency (crime laboratory), or like agency, in the United States. This included active, retired or contracted latent print examiners as eligible candidates to participate in this study.

D. Accessible Population

Accessibility was limited to latent print examiners for whom the MDPD FSB was able to obtain an e-mail address by querying the membership of the International Association for Identification (IAI), as well as latent print examiners who volunteered to participate in this study.

E. Sampling Plan & Setting

The sampling plan for this study utilized an abstract population. Active and retired latent print examiners in the United States with a functional email address who were a member of the International Association for Identification (IAI) were invited to participate in this research. The IAI list identified participants whose discipline was in latent fingerprint examination. Applications were also made available to any qualified latent print examiner, regardless of affiliation with a professional organization. The accessible population included approximately 1,700 latent print examiners in the United States. To ensure confidentiality, the researchers at the MDPD FSB invited latent print examiners to respond via email or fax.

F. Instrumentation

This study utilized two methods of instrumentation: a questionnaire that included the participant’s qualifications and demographics, as well as answer sheets for each of the three (3) phases of experimental exercises. The questionnaire took less than ten (10) minutes to complete. The experimental exercises took each participant approximately 30-35 hours to complete in Phase 1, 20-25 hours to complete in Phase 2, and 15-20 hours to complete in Phase 3. These approximate time frames were based on the length of time necessary for participants from the MDPD FSB to complete the exercises.

The experimental exercises conducted to test ACE and ACE-V methodologies were similar to those devised and utilized by Langenburg (2009). His pilot study consisted of six (6) latent print examiners at the Minnesota Bureau of Criminal Apprehension, St. Paul, Minnesota. Using this same concept, the researchers also increased the number of participants, included latent print examiners across multiple laboratories/agencies, and increased the number of ACE and ACE-V examinations conducted.

G. Data Collection Methods

The researchers performed the following steps:

1. Received National Institute of Justice (NIJ) approval.
2. Prepared an announcement and questionnaire for dissemination.
3. Emailed the announcement and questionnaire to the IAI membership whose discipline was listed as Latent Fingerprint Examination.

4. Created a Microsoft Access database to record participant qualifications, demographics and mailing information.
5. Identified 13 volunteers to create latent prints and fingerprint and palm print standards.
6. Collected 2,711 latent prints, and 13 fingerprint and palm print standards from these volunteers.
7. Designated 3 of the 13 fingerprint and palm print standards as sources that would not be utilized for comparison during testing.
8. Evaluated 320 latent prints of varying quality and quantity of information and assigned ratings from strength of value, latent in agreement with the standards, and difficulty of comparison.
9. Selected 160 latent prints based on similar difficulty ratings that were evaluated by the IAI certified latent print examiners.
10. Divided and designated the 160 latent prints into two groups (source present and source not present). There were 112 latent prints (70%) that were selected where the source was present and 48 latent prints (30%) that were selected where the source was not present.
11. Divided and designated the latent prints based on the range of difficulty ratings for both source present and source not present groups. For the group where the source was present, there were 28 latent prints (25%) that were rated Insufficient to Difficult, 42 latent prints (37.5%) that were rated Difficult to Moderate, and 42 latent prints (37.5%) that were rated Moderate to Easy. For the group where the source was not present there were 12 latent prints (25%) that were rated in the Insufficient to Difficult range, 18 latent prints (37.5%) that were rated Difficult to Moderate, and 18 latent prints (37.5%) that were rated Moderate to Easy.
12. Divided the 160 latent prints using a random number generator to comprise the final 80 latent prints used for this research. For the group where the source was present, there were 14 latent

prints (25%) that were rated Insufficient to Difficult, 21 latent prints (37.5%) that were rated Difficult to Moderate, and 21 latent prints (37.5%) that were rated Moderate to Easy. For the group where the source was not present there were 6 latent prints (25%) that were rated in the Insufficient to Difficult range, 9 latent prints (37.5%) that were rated Difficult to Moderate, and 9 latent prints (37.5%) that were rated Moderate to Easy.

13. Created answer sheets and instructions for the participants.
14. Prepared and mailed 140 identical Phase 1 test packets to the participants which included the following items:
 - 40 latent prints
 - 10 sets of known fingerprint and palm print standards
 - 1 answer sheet, including instructions
 - 1 return envelope
15. Instructed the participants to conduct 40 ACE examinations and compare each latent print to 3 specific standards of the 10 standards that were provided. The instructions directed the participants to return all testing materials and their answer sheets to the researchers via mail.
16. Recorded and analyzed the data from 109 Phase 1 test packets utilizing Microsoft Excel for 16 weeks.
17. Prepared and mailed 109 identical Phase 2 test packets to the participants in two randomly created groups (A, B) which included the following items:
 - 20 latent prints (Group A)
 - 20 latent prints (Group B)
 - 10 sets of known fingerprint and palm print standards to each group
 - 1 answer sheet, including instructions to each group

- 1 return envelope to each group
18. Instructed the participants to conduct 20 ACE examinations and compare each latent print to 3 specific standards of the 10 standards that were provided. The instructions directed the participants to return all testing materials and their answer sheet to the researchers via mail.
19. Recorded and analyzed the data from 88 Phase 2 test packets utilizing Microsoft Excel for 12 weeks.
20. Prepared and mailed 88 individual Phase 3 test packets to the participants which included the following items.
- A selected amount of latent trials based on 52 latent prints (Subgroup 1)
 - A selected amount of latent trials based on 61 latent prints (Subgroup 2)
 - 10 sets of fingerprint and palm print standards to each subgroup
 - 1 answer sheet, including instruction to each subgroup
 - 1 return envelope to each subgroup
 - 1 participant exit questionnaire (See Appendix A)
21. Instructed the participants in Subgroup 1 to conduct verifications and Subgroup 2 to conduct second verifications of the initial identification, and compare each latent print to 1 specific standard of the 10 standards that were provided. The instructions directed the participants to return all testing materials and their answer sheets to the researchers via mail.
22. Recorded and analyzed the data from 84 Phase 3 test packets utilizing Microsoft Excel for 16 weeks.
23. Presented the data to a professor from the Department of Statistics at Florida International University for statistical analyses.

H. Data Coding

The data coding for the participants and questionnaire is described below:

1. Each participant was assigned a number 1 to end.
2. Type of Law Enforcement Agency employed by was coded as: Other (0), Local (1), State (2), and Federal (3)
3. Group A was coded as (1) and Group B was coded as (2)
4. Subgroup 1 was coded as (1) and Subgroup 2 was coded as (2)
5. Sworn was coded as: yes (1), no (2)
6. Employed as a latent print examiner by a Law Enforcement Agency was coded as: yes (1), no (2)
7. Latent Print Examinations as a Primary Duty was coded: yes (1), no (2)
8. Retired latent print examiner from a law enforcement agency was coded as: yes (1), no (2)
9. Examine other types of evidence in addition to latent fingerprints was coded as: yes (1), no (2)
10. Independent Contractor performing latent print examinations was coded as: yes (1), no (2)
11. Member of the IAI was coded as: yes (1), no (2)
12. IAI certified latent print examiner was coded as: yes (1), no (2)
13. Apply ACE-V methodology in latent fingerprint examinations was coded as: yes (1), no (2)
14. High School Degree was coded as: yes (1), no (2)
15. College Degree was coded as: yes (1), no (2)
16. Science Degree was coded as: yes (1), no (2)
17. Duration of a structured latent fingerprint training program was coded as: 0-1 years (1), 1-2 years (2), more than 2 years (3)

18. Years of active casework experience completed was coded as: >15 years (1), 10-15 years (2), 5-10 years (3), and 1-5 years (4)

The data coding for the latent prints is described below:

1. 80 latent prints were each assigned a unique label comprised of a Greek Letter for their source, lift number created from that source, a lower case letter for impression on that lift, and if applicable a number indicating which image of that impression was used.
2. Strength of Value of Latent Print:
 - Insufficient to Difficult (0-7) was coded as (1)
 - Difficult to Moderate (8-14) was coded as (2)
 - Moderate to Easy (15-21) was coded as (3)
3. Difficulty of Comparison:
 - Source not Present was coded as (0)
 - Insufficient to Difficult (0-7) was coded as (1)
 - Difficult to Moderate (8-14) was coded as (2)
 - Moderate to Easy (15-21) was coded as (3)

The data coding for the Phase 1 answer sheets is described below:

1. Analysis
 - Clarity was coded as: Blank response (0), Maximum level of detail observed Level 1 (1), Maximum level of detail observed Level 2 (2), and Maximum level of detail observed Level 3 (3)
 - Anatomical source was coded as: Fingerprint (1), Palm Print (2), Impression (3)
 - Orientation was coded as: Certain (1), Uncertain (2)
 - Sufficiency determination was coded as: Value (1), No Value (2)

2. Conclusions

- Identification was coded as: Correct (1), Incorrect (2)
- Exclusion was coded as: Correct (3), Incorrect (4)
- Non-Identification was coded as: Exclusion (1), Inconclusive (2)

The data coding for the Phase 2 answer sheets is described below:

1. Analysis

- Clarity was coded as: Blank response (0), Maximum level of detail observed Level 1 (1), Maximum level of detail observed Level 2 (2), and Maximum level of detail observed Level 3 (3)
- Anatomical source was coded as: Fingerprint (1), Palm Print (2), Impression (3)
- Orientation was coded as: Certain (1), Uncertain (2)
- Sufficiency determination was coded as: Value (1), No Value (2)

2. Conclusions

- Identification was coded as: Correct (1), Incorrect (2)
- Exclusion was coded as: Correct (3), Incorrect (4)
- Non-Identification was coded as: Exclusion (1), Inconclusive (2)

The data coding for the Phase 3 answer sheets is described below:

1. Verification Conclusions

- Verify Correct Identification (Agree) was coded as (16)
- Verify Incorrect Identification (Agree) was coded as (17)
- Verify Correct Identification (Disagree) was coded as (18)
- Verify Incorrect Identification (Disagree) was coded as (19)
- Inconclusive (Source Present) was coded as (20)

- Inconclusive (Source Not Present) was coded as (23)

2. Repeat Conclusions

- Repeat Correct Identification (Agree) was coded as (1)
- Repeat Incorrect Identification (Agree) was coded as (2)
- Repeat Incorrect Exclusion (Disagree) was coded as (21)
- Does Not Repeat Correct Identification (Disagree) was coded as (3)
- Does Not Repeat Correct Identification (Inconclusive) was coded as (4)
- Does Not Repeat Incorrect Identification (Disagree) was coded as (5)
- Does Not Repeat Incorrect Identification (Inconclusive) was coded as (6)
- Repeat Inconclusive was coded as (7)
- Does Not Repeat Inconclusive (Agree with Correct Identification) was coded as (8)
- Does Not Repeat Inconclusive (Disagree with Correct Identification) was coded as (9)
- Does Not Repeat Incorrect Exclusion (Agree with Correct Identification) was coded as (14)
- Does Not Repeat Incorrect Exclusion (Inconclusive) was coded as (22)

3. Verification Conclusions (Bias)

- Verify Correct Identification (Agree) was coded as (16)
- Verify Incorrect Identification (Agree) was coded as (17)
- Verify Correct Identification (Disagree) was coded as (18)
- Verify Incorrect Identification (Disagree) was coded as (19)
- Inconclusive (Source Present) was coded as (20)

- Inconclusive (Source Not Present) was coded as (23)

4. Repeat Conclusions (Bias)

- Repeat Correct Identification (Agree) was coded as (1)
- Repeat Incorrect Identification (Agree) was coded as (2)
- Repeat Incorrect Exclusion (Disagree) was coded as (21)
- Does Not Repeat Correct Identification (Disagree) was coded as (3)
- Does Not Repeat Correct Identification (Inconclusive) was coded as (4)
- Does Not Repeat Incorrect Identification (Disagree) was coded as (5)
- Does Not Repeat Incorrect Identification (Inconclusive) was coded as (6)
- Repeat Inconclusive was coded as (7)
- Does Not Repeat Inconclusive (Agree with Correct Identification) was coded as (8)
- Does Not Repeat Inconclusive (Disagree with Correct Identification) was coded as (9)
- Does Not Repeat Incorrect Exclusion (Agree with Correct Identification) was coded as (14)
- Does Not Repeat Incorrect Exclusion (Disagree with Correct Identification) was coded as (15)
- Does Not Repeat Incorrect Exclusion (Inconclusive) was coded as (22)

I. Descriptive Analysis

Descriptive analysis was used to characterize the participants. Descriptive analysis included type of government agency, number of years of latent print examination experience, current work status, and if they were an IAI certified latent print examiner.

J. Data Analysis Methods

Simple descriptive scores were used to analyze all variables. Statistical analysis was performed utilizing Microsoft Excel to answer the five research questions. An independent statistician performed a statistical analysis from the data generated.

K. Definitions

For this research study, the following definitions apply:

1. *Accuracy* – The ability of the participant to correctly identify or exclude a latent print to a known source(s).
2. *Bias* – The ability of the participant to reproduce or repeat a conclusion when presented with two previous conclusions and asked to conduct a second verification.
3. *Exclusion* - The determination by a participant that there is sufficient quality and quantity of detail in disagreement to conclude that two areas of friction ridge impressions did not originate from the same source.
4. *False Negative Discovery Rate (FNDR)* - Percentage of the time the participant made an erroneous exclusion when reporting the categorical opinion of an Exclusion.
5. *False Negative Rate (FNR)* - Percentage of the time the participant made an erroneous exclusion when given the possibility of making any of the three categorical opinions (Identification, Inconclusive, and Exclusion)
6. *False Positive Discovery Rate (FPDR)* - Percentage of the time the participant made an erroneous identification when reporting the categorical opinion of an Identification.

7. *False Positive Rate (FPR)* – Percentage of the time the participant made an erroneous identification when given the possibility of making any of the three categorical opinions (Identification, Inconclusive, and Exclusion).
8. *Identification* - The determination by a participant that there is sufficient quality and quantity of detail in agreement to conclude that two areas of friction ridge impressions did originate from the same source.
9. *Inconclusive* - The determination by a participant that there is neither sufficient agreement to individualize, nor sufficient disagreement to exclude. Participants were asked to record their reason for inconclusive results, which included the following choices: “Poor quality standards,” “Sufficient detail for comparison, but insufficient to identify,” “Cannot fully exclude all three individuals,” or “Other.”
10. *Negative Predictive Value (NPV)* - Percentage of the time the participant made a correct exclusion when reporting the categorical opinion of an Exclusion.
11. *No Value* - The determination by a participant that the quantity and quality of information present in the latent print is not sufficient for identification. Impressions that are insufficient for identification, but sufficient for exclusion were, for the purposes of this test, deemed “No Value”.
12. *Positive Predictive Value (PPV)* - Percentage of the time the participant made a correct identification when reporting the categorical opinion of an Identification.
13. *Precision* – The ability of the participant(s) to reproduce and repeat the same conclusion.
14. *Repeatability* – The ability of a participant to provide the same conclusion upon re-evaluation of the same latent print.

15. *Reproducibility* - The ability of multiple participants to examine the same latent print and reach the same conclusions independently.
16. *Second Verification* – the determination by a participant of agreement or disagreement with an initial conclusion when provided with a verification conclusion.
17. *Value* - The determination by a participant that the quantity and quality of information present in the latent print is sufficient for identification.
18. *Verification* – the determination by a participant of agreement or disagreement with an initial conclusion.

L. Internal Validity Strengths

1. The internal validity of the quantitative data was valid due to the procedures used to assemble the tests.
2. All the test materials were assembled in a crime laboratory setting.
3. All known standards and unknown latent impression were labeled with a letter (known standard) or unique label (unknown latent impressions).
4. Packets were used to separate materials for all three phases of the test.
5. Three IAI certified latent print examiners with a combined experience of over 65 years evaluated every latent impression and assigned difficulty ratings that were used for testing materials.
6. The researchers at the MDPD FSB printed all testing materials from their original electronic form and evaluated each of the participants testing materials to ensure they maintained the same quality and quantity of information.

7. The participants had to conduct a “search” of different known fingerprint and palm print standards and were not given a 1:1 comparison during two phases of testing.
8. The latent prints utilized during testing were rated according to their level of difficulty.

M. Internal Validity Weaknesses

1. The validity of this study was dependent upon the accuracy of assembling the tests.
2. Communication between participants about test materials may have threatened the internal validity.
3. During some of the testing, participants were presented with latent print comparisons with prior conclusions which could have affected results.

N. External Validity Strengths

1. The participants all had one year of active casework experience conducting latent print examinations.
2. The participants represented a large sample size across multiple laboratories/agencies nationwide.

O. External Validity Weaknesses

1. The researchers presumed the participants followed ACE methodology as specified by SWGFAST.
2. The researchers had no control over the equipment used by the participants.
3. The training and experience of the participants could have been an external weakness.

4. The participants may have remembered their own prior conclusions for latent prints sent for verification testing (reproducibility and repeatability).

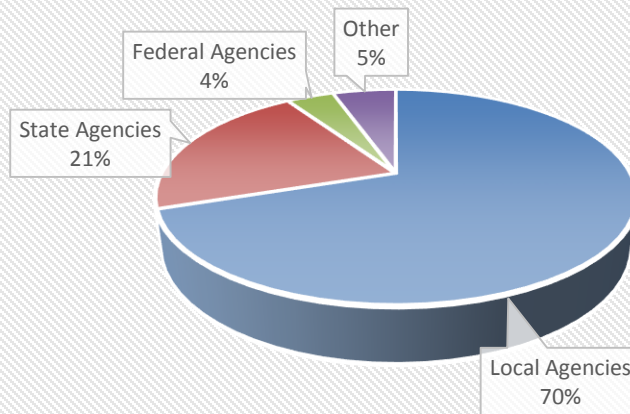
IV. RESULTS

In this section, participant demographics and data from this study are discussed to evaluate accuracy and precision in latent print examination decisions. Data was evaluated based on the overall results, analysis of the latent print, difficulty of the latent print comparison, as well as the qualifications and demographical information provided by the participants.

A. Participant Demographics

The participants in Phases 1, 2, and 3 of this study represented a total of 76 different law enforcement agencies across the United States. Of the 109 participants who participated in this study, 71 of the participants were from 53 different local agencies, 29 of the participants were from sixteen different state agencies, and five of the participants were from three different federal agencies (Figure 7). Three of the participants were private contractors and one participant was retired. The participants also varied in latent print examination experience, IAI latent print certification, and length of a structured latent fingerprint training program (Figures 8, 9, 10, and 11).

**Figure 7: Law Enforcement Agencies Represented
(N = 109 Participants)**



**Figure 8: Latent Print Examination Experience
(N = 109 Participants)**

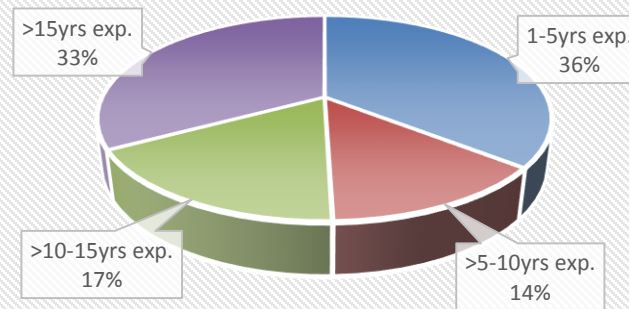


Figure 9: IAI Latent Print Certification (N =109 Participants)

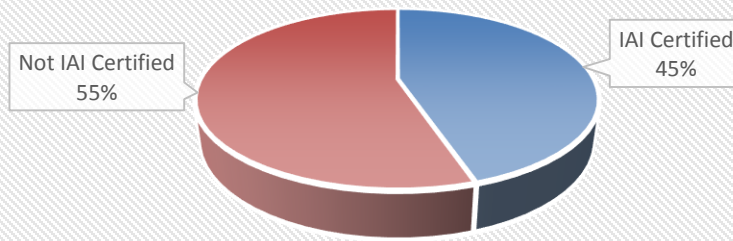


Figure 10: Completion of a Structured Latent Fingerprint Training Program (N =109 participants)

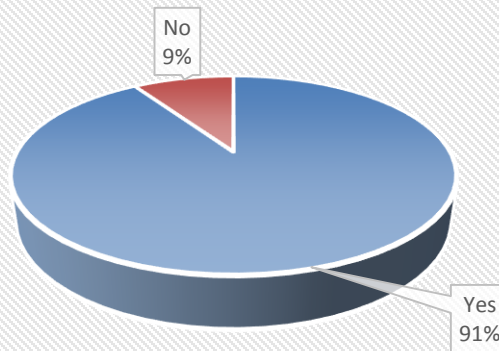
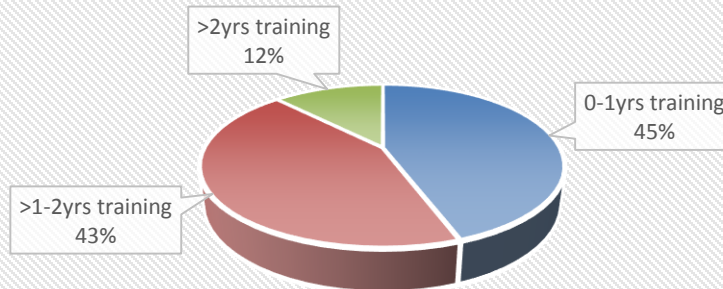


Figure 11: Years Completed in a Structured Latent Fingerprint Training Program (N =99 Participants)



B. Sufficiency Determinations

The participants reported a total of 5,963 sufficiency determinations. Not all participants completed every ACE trial presented to them in this study. The results of sufficiency determinations for Phases 1 and 2 are shown in Table 2.

Table 2: Sufficiency Determinations (109 Participants)			
Sufficiency Determination	Phase 1 109 Participants	Phase 2 88 Participants	Total Decisions
Value	3,210	1,342	4,552
No Value	1,023	388	1,411
Total Decisions	4,233	1,730	5,963

Table 3 represents the results of sufficiency determinations based on the *Strength of Value* of the latent print.

Table 3: Strength of Value Sufficiency Determinations (109 Participants) (N =5,963 Decisions)			
<i>Strength of Value</i> Rating	Phase 1 109 Participants	Phase 2 88 Participants	Total Decisions
Value (0-7)	109	85	194
No Value (0-7)	840	309	1,149
Value (>7-14)	1,296	567	1,863
No Value (>7-14)	175	76	251
Value (>14-21)	1,805	690	2,495
No Value (>14-21)	8	3	11
Total Decisions	4,233	1,730	5,963

C. Accuracy – ACE and ACE-V

Accuracy was measured in terms of the participant’s ability to correctly identify or exclude latent prints to known standards using both the ACE and ACE-V methodologies. The accuracy of ACE and ACE-V examinations are reported as an overall participant error rate after participants

made a sufficiency determination that a latent was of “value” for Identification. As with sufficiency determinations, not all the participants completed every ACE and ACE-V trial presented to them. Error rates are reported with consideration to inconclusive decisions.

Additionally, the accuracy of ACE examinations are also reported according to the participant’s years of latent print examination experience, IAI certification, and the *Difficulty of Comparison* rating when the source was present that were assigned during ACE trials.

Error Rates of ACE Trials

For participant performance related to ACE accuracy (Q1), the combined results of three categorical opinions (i.e., Identification, Exclusion, and Inconclusive) from Phase 1 and 2 ACE trials were evaluated. The results of ACE examinations for Phases 1 and 2 are shown in Table 4 and the ACE error rates for Phases 1 and 2 are shown in Table 5.

Table 4: ACE Examinations			
	Same Source	Different Source	Totals
Identification	2,457	42	2,499
Inconclusive	446	403	849
Exclusion	235	953	1,188
Totals	3,138	1,398	4,536

Note: The number of erroneous Identifications and Exclusions are in bold.

Table 5: Error Rates for ACE Examinations		
	With Inconclusives	Without Inconclusives
False Positive Rate	3.0%	4.2%
False Negative Rate	7.5%	8.7%
False Positive Discovery Rate (FPDR)	1.6%	
* False Negative Discovery Rate (FNDR)	7.6%	
Positive Predictive Value (PPV)	98.3%	
Negative Predictive Value (NPV)	92.4%	

**Note: In calculating the False Negative Discovery Rate and Negative Predictive Value, consideration was given to the number of standards presented to the participant.*

Error Rates: ACE Trials and Latent Fingerprint Examination Experience

The results of Phase 1 and 2 ACE examinations are shown in Table 6 according to years of latent fingerprint examination experience.

Table 6: ACE Trials & Error Rates (With Inconclusives) Latent Print Examination Experience									
Years of Exp.	# of Participants	Correct Identification	Incorrect Identification	Inconclusive	Correct Exclusions	Incorrect Exclusions	Totals	FPR	FNR
1-5	39	852	21	357	286	84	1,600	4.3%	7.6%
>5-10	15	318	6	89	127	38	578	3.4%	9.5%
>10-15	19	435	5	124	191	48	803	2.0%	8.8%
+15	36	852	10	279	349	65	1,555	2.1%	6.0%
Totals	109	2,457	42	849	953	235	4,536		

Error Rates: ACE Trials and IAI Certification

The results of ACE examinations for Phases 1 and 2 by IAI Certification are shown in Table 7.

Table 7: ACE Trials & Error Rates (With Inconclusives) IAI Latent Print Certification									
IAI Cert.	# of Participants	Correct Identification	Incorrect Identification	Inconclusive	Correct Exclusions	Incorrect Exclusions	Totals	FPR	FNR
Yes	49	1,173	19	315	511	113	2,131	2.8%	7.7%
No	60	1,284	23	534	442	122	2,405	3.2%	7.7%
Totals	109	2,457	42	849	953	235	4,536		

Error Rates: ACE Trials and Difficulty of Comparison Rating

The results of ACE examinations for Phases 1 and 2 by *Difficulty of Comparison* Rating are shown in Table 8. The FPR could not be reported since the source was present for *Difficulty of Comparison* trials. Therefore, the FPDR and FNR are reported.

Table 8: ACE Trials & Error Rates (With Inconclusives)							
<i>Difficulty of Comparison Rating</i>							
Difficulty of Comparison Rating	Correct Identification	Incorrect Identification	Inconclusive	Incorrect Exclusions	Totals	FPDR	FNR
0-7	40	0	175	14	229	0.0%	6.1%
>7-14	1,001	20	233	172	1,426	2.0%	12.2%
>14-21	1,416	19	38	49	1,522	1.3%	3.3%
Totals	2,457	39	446	235	3,177		

Participant Error Rates (ACE-V)

For participant performance related to ACE-V accuracy (Q2), the combined verification results from Phase 3 were evaluated. The results of Phase 3 ACE-V examinations are shown in Table 9, and the Phase 3 ACE-V error rates are shown in Table 10.

Table 9: ACE-V Examinations			
	Same Source	Different Source	Totals
Identification	487	0	487
Inconclusive	15	2	17
Exclusion	15	13	28
Totals	517	15	532

Note: The number of erroneous Identifications and Exclusions are in bold.

Table 10: Error Rates for ACE-V Examinations		
	With Inconclusives	Without Inconclusives
False Positive Rate	0.0%	0.00%
False Negative Rate	2.9%	3.0%
False Positive Discovery Rate (FPDR)	0.0%	
* False Negative Discovery Rate (FNDR)	53.6%	
Positive Predictive Value (PPV)	100.%	
Negative Predictive Value (NPV)	46.4	

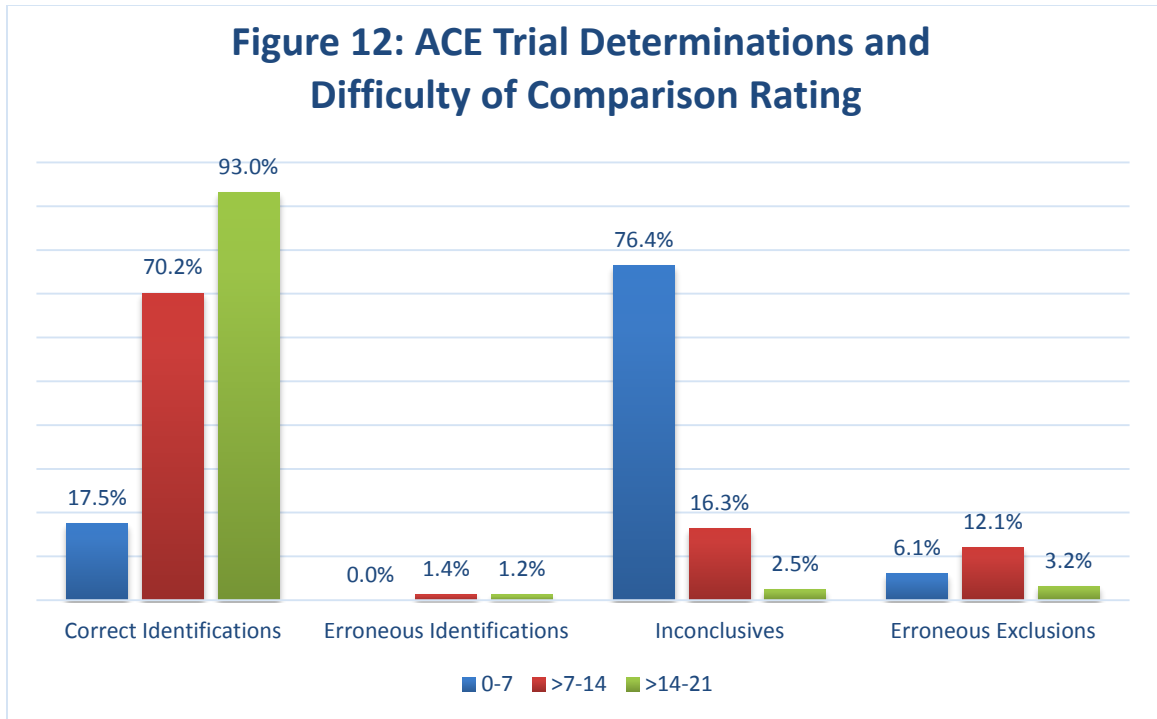
D. Significantly Varied Results

To determine if the participants would reach significantly varied results using the ACE methodology (Q3), the combined results of three categorical opinions (Identification, Exclusion, and Inconclusive) from Phase 1 and 2 ACE trials were grouped by their *Difficulty of Comparison*

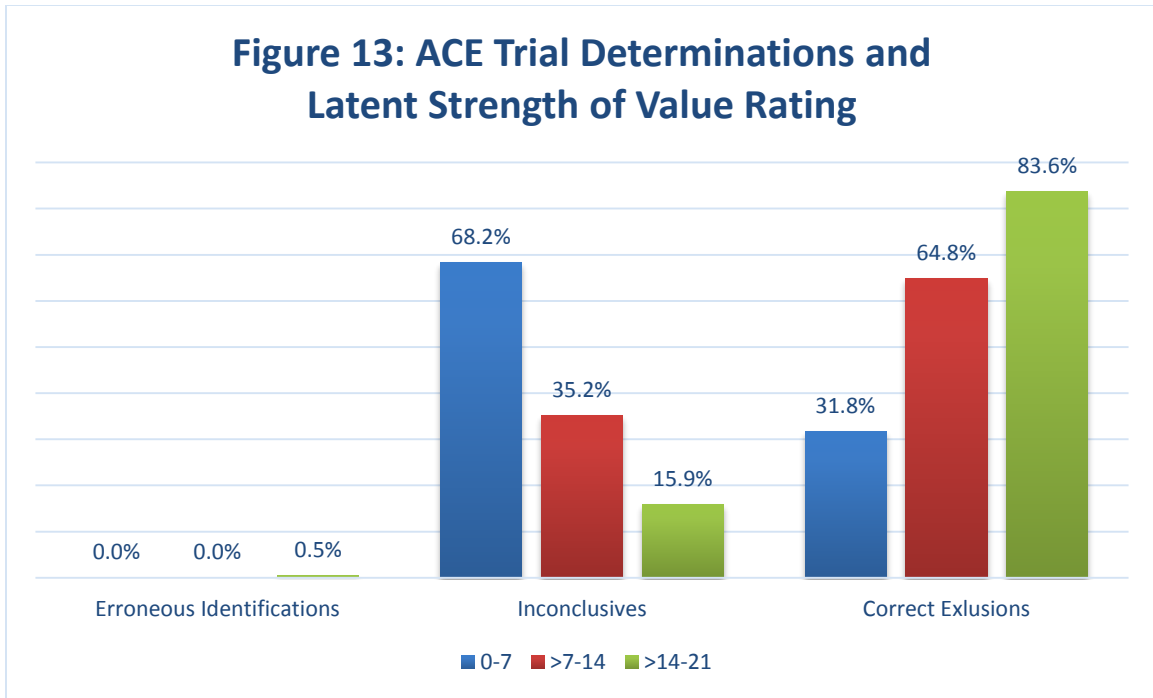
(source present) and *Strength of Value* (source not present) ratings. The results of ACE trials from Phase 1 and 2 are shown in Table 11.

Table 11: ACE Trials Participant Decisions							
Source Present (Y/N)	# of Latent Prints	# of Decisions	Correct Identifications	Erroneous Identifications	Inconclusives	Correct Exclusions	Erroneous Exclusions
Yes	56	3,177	2,457	39	446	N/A	235
No	24	1,359	N/A	3	403	953	N/A
Totals	80	4,536	2,457	42	849	953	235

For Phase 1 and 2 ACE trials when the source was present (N=56), there were 229 conclusions: 40 identifications; no erroneous identifications; 175 inconclusives; and fourteen erroneous exclusions for latents with a *Difficulty of Comparison* rating of Insufficient to Difficult (0-7). For latents with a *Difficulty of Comparison* rating of Difficult to Moderate (>7-14), there were 1,426 conclusions: 1,001 identifications; 20 erroneous identifications; 233 inconclusives; and 172 erroneous exclusions. For latents with a *Difficulty of Comparison* rating of Moderate to Easy (>14-21), there were 1,522 conclusions: 1,416 identifications; 19 erroneous identifications; 38 inconclusives; and 49 erroneous exclusions. A comparison of ACE trial determinations based on *Difficulty of Comparison* ratings is shown in Figure 12.



For Phase 1 and 2 ACE trials when the source was not present (N=24), there were 129 conclusions: 0 erroneous identifications; 88 inconclusives; and 41 correct exclusions for latents with a *Strength of Value* rating of Insufficient to Difficult (0-7). For latents with a *Strength of Value* rating of Difficult to Moderate (>7-14), there were 620 conclusions: no erroneous identifications; 218 inconclusives; and 402 correct exclusions. For latents with a *Strength of Value* rating of Moderate to Easy (>14-21), there were 610 conclusions: three erroneous identifications; 97 inconclusives; and 510 correct exclusions. A comparison of ACE trial determinations based on *Strength of Value* ratings is shown in Figure 13.



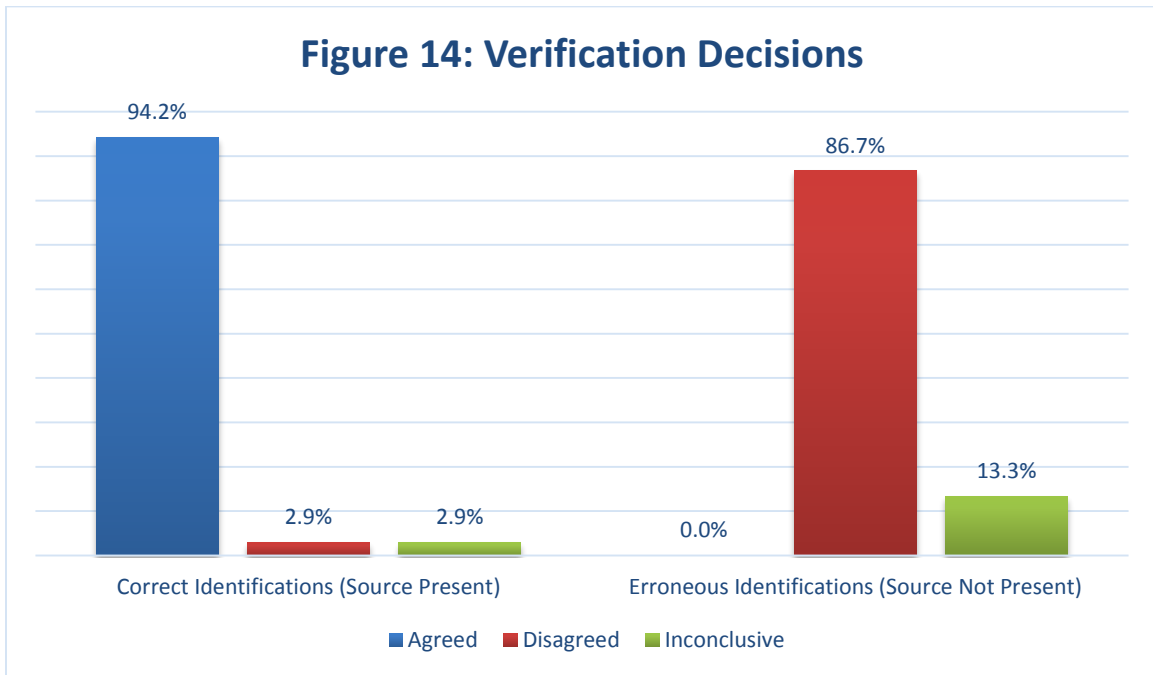
E. Precision

Precision was measured in terms of the participant’s ability to reproduce and repeat the same conclusion after participants made sufficiency determinations that a latent was of “Value” for identification.

To determine if the participants would reproduce conclusions from comparisons of unknown latent prints to known standards made by other participants using the ACE methodology (Q4), the results of identification decisions from Phase 2 were sent to different participants in Phase 3 in order to determine if they would agree, disagree, or came to an inconclusive decision.

The number of latent prints presented to participants for verification were based on 25 latent prints for a total of 532 participant verification decisions: 517 verification decisions when the source was present and fifteen verification decisions when the source was not present. When the source was present and the participant was presented with a correct identification, a second participant agreed with the correct identification 487 times, disagreed fifteen times, and came to

an inconclusive decision fifteen times. In addition, when the source was not present and the participant was presented with an erroneous identification, a second participant never agreed with the incorrect identification, disagreed thirteen times and came to an inconclusive two times. Figure 14 compares determinations made when the source was present and not present.

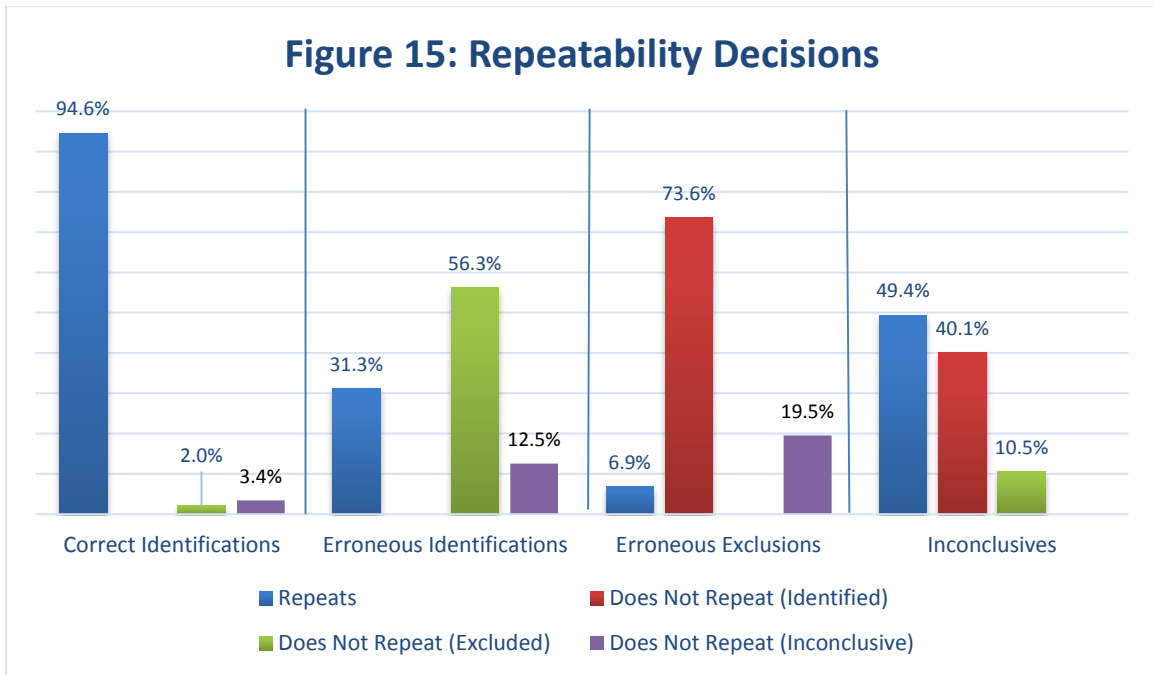


Note: The average Strength of Value rating for the 25 Latent Prints was 14.15.

To determine if the participants would repeat their own conclusions from comparisons of unknown latent prints to known standards using the ACE methodology (Q4), the results of identification decisions, erroneous exclusions, and inconclusive results (where the source was present) from Phase 1 were sent to the same participants in Phase 3.

The number of latent prints presented to participants for repeatability was based on 27 latent prints for a total of 1,311 participant decisions. When presented with their previous correct identification, participants repeated their answer 980 times and did not repeat their answer 56 times; participants incorrectly excluded the source 21 times and reported an inconclusive decision

35 times. When presented with their previous incorrect identification, participants repeated their answer five times and did not repeat their answer eleven times. Participants correctly excluded the source nine times and reported an inconclusive decision two times. When presented with their previous incorrect exclusion and given the correct source, participants repeated their answer six times and did not repeat their answer 81 times. Participants correctly identified the source 64 times and reported an inconclusive decision seventeen times. When presented with the correct source as a verification for their previously reported inconclusive decision, participants repeated their answer 85 times and did not repeat their answer 87 times. Participants correctly identified the source 69 times and incorrectly excluded the source eighteen times. Figure 15 compares determinations made that were repeated and not repeated.



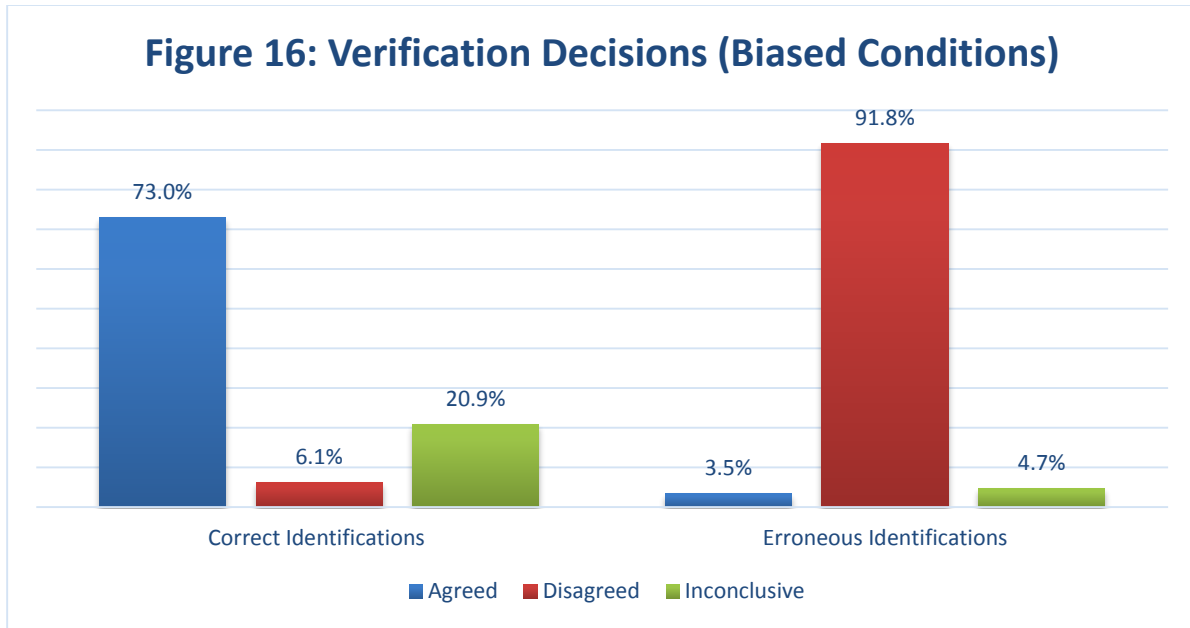
Note: The average Strength of Value rating for the 27 Latent Prints was 12.38.

F. Bias

Bias was measured in terms of the participant’s ability to reproduce and repeat the same conclusion when asked to verify an identification that included a previous verification conclusion (second verification).

To determine if the participants would reproduce conclusions from comparisons of unknown latent prints to known standards made by other participants using the ACE methodology under biased conditions (Q5), the results of identification decisions from Phase 2 were sent to different participants in Phase 3 in order to determine if they would agree, disagree, or come to an inconclusive decision.

The number of latent prints presented to participants for verification under biased conditions was based on 37 latent prints for a total of 329 participant verification decisions: 244 verification decisions when the source was present and 85 verification decisions when the source was not present. When the source was present and the participant was presented with a correct identification, a second participant agreed with the correct identification 178 times, disagreed fifteen times, and came to an inconclusive decision 51 times. In addition, when the source was not present and the participant was presented with an erroneous identification, a second participant agreed with the incorrect identification three times, disagreed 78 times and came to an inconclusive decision four times. The comparison of determinations made under biased conditions when the source was present and not present is shown in Figure 16.

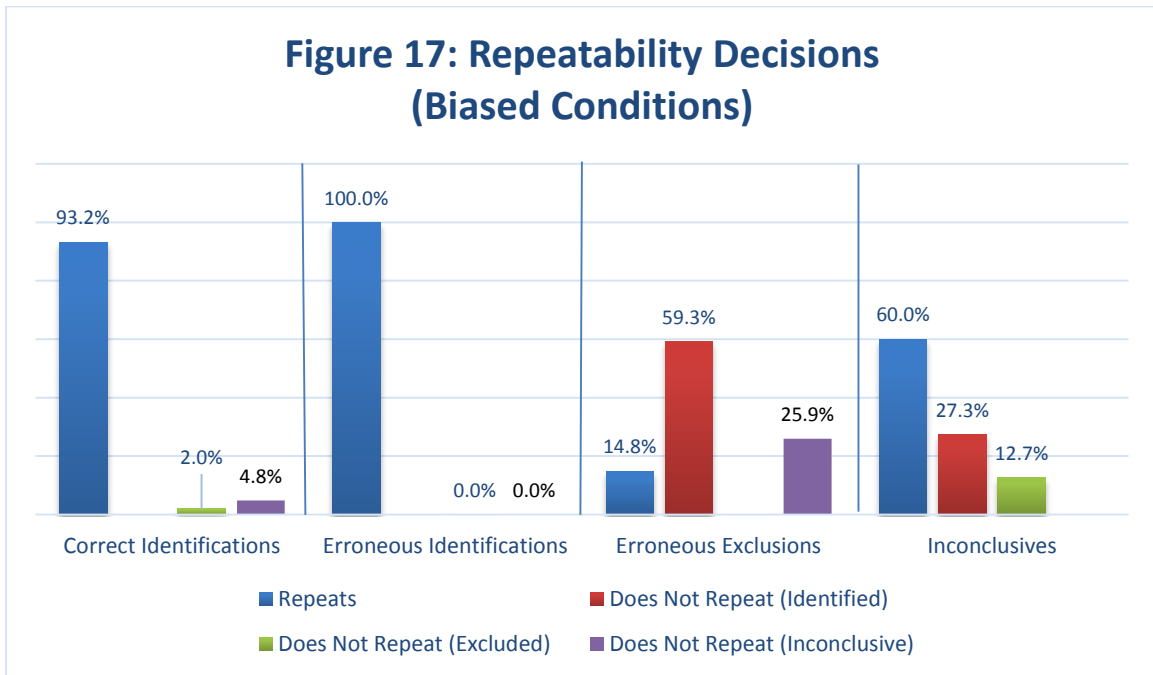


Note: The average Strength of Value rating for the 37 Latent Prints was 13.85.

To determine if the participants would repeat their own conclusions from comparisons of unknown latent prints to known standards using the ACE methodology under biased conditions (Q5), the results of identification decisions, erroneous exclusions, and inconclusive results (where the source was present) from Phase 1 were sent to the same participants in Phase 3.

The number of latent prints presented to participants for repeatability under biased conditions was based on 24 latent prints for a total of 333 participant decisions. When presented with their own previous correct identification, participants repeated their answer 233 times and did not repeat their answer seventeen times; the participants incorrectly excluded the source five times and came to an inconclusive decision twelve times. A single participant repeated their erroneous identification. When presented with their own previous incorrect exclusion and given the correct source, participants repeated their answer four times and did not repeat their answer 23 times; participants correctly identified the source sixteen times and came to an inconclusive decision on seven occasions. When presented with the correct source as a verification for their previously

reported inconclusive decision, participants repeated their answer 33 times and did not repeat their answer 22 times; participants correctly identified the source fifteen times and incorrectly excluded the source seven times. Figure 17 compares determinations made under biased conditions that were repeated and not repeated.



Note: The average Strength of Value rating for the 24 Latent Prints was 14.00.

V. CONCLUSIONS

A. Discussion and Findings

Sufficiency Determinations

To determine how the quantity and quality of latent prints affects decision making abilities in latent print examinations, the authors created a *Strength of Value* rating scale based on three factors; number of minutiae present, number of minutiae formations present, and clarity (deposition pressure) present for each latent print. Of the 5,963 sufficiency determinations that

were observed, there were 4,552 (76.3%) of value decisions and 1,411 (23.6%) no value decisions reported. Additionally, of the 20 latents that were rated in the Insufficient to Difficult (0-7) category, participants reported an of value for identification decision 14.5% of the time and a no value decision 85.6% of the time. Of the 30 latents in the Difficult to Moderate (>7-14) range, participants reported an of value for identification decision 88.1% of the time and a no value decision 11.9% of the time. Of the 30 latents in the Moderate to Easy (>14-21) range, participants reported an of value for identification decision 99.6% of the time and a no value decision 0.4% of the time.

Erroneous Identifications of ACE Trials

Of the 42 erroneous identifications that were reported during ACE trials, 28 of 109 participants committed an identification error. There were nineteen participants who committed one error; six participants who committed two errors; two participants who committed three errors; and one participant who committed five errors. The identification errors occurred on 21 of 80 different latents used for this research. There were nine latents with one reported identification error; five latents with two reported identification errors; five latents with three reported identification errors; and two latents with four reported identification errors.

In assessing these errors, it was noted that in 35 of the 42 erroneous identifications the participants appear to have made a clerical error, but the authors could not determine this with certainty. A clerical error was defined as a circumstance in which the participant chose the correct standard from the three standards presented, however, the opposite finger (i.e., Left Index Finger as opposed to Right Index Finger), opposite palm (i.e., Left Palm as opposed to Right Palm) or incorrect finger (i.e., Left Index Finger as opposed to Left Ring Finger) was reported. In addition, a clerical error may have occurred in which the incorrect standard was chosen, but the correct

finger or palm was reported. The remaining seven errors appear to be true erroneous identifications, in which the incorrect standard was reported, or where the source was not present for that particular trial.

In investigating the seven erroneous identifications further, one examiner committed three of these errors, another examiner committed two of these errors, and the remaining two errors were committed by two separate examiners. In addition, the seven erroneous identifications occurred on four separate latent trials (two source present and two source not present) with an average *Difficulty of Comparison* rating of 10.6 and an average *Strength of Value* rating of 13.2 (See Appendix B). The four participants varied in terms of their experience, certification, and duration of a latent fingerprint training program.

Erroneous Identifications vs Erroneous Exclusions of ACE & ACE-V Trials

In comparing the number of erroneous identifications and erroneous exclusions for ACE & ACE-V trials, the error rate was less for ACE-V trials. In comparing the number of reported erroneous identifications to erroneous exclusions, the exclusion error rate was higher in both ACE and ACE-V trials. Other fingerprint research studies have reported similar findings.

Error Rates: ACE Trials and Demographics

In evaluating identification error rates within this study as it relates to participant demographics, the data indicates an identification error rate decrease for participants with more latent print examination experience. However, identification error rates were nearly the same for participants with or without IAI latent print certification.

In evaluating exclusion error rates within this study as it relates to participant demographics, the data indicates that exclusion error rates were higher than identification error

rates irrespective of the participant’s years of latent print examination experience. Additionally, exclusion error rates did not change for participants with or without IAI latent print certification.

Error Rates: ACE Trials and Difficulty of Comparison Rating

There was a significant difference in exclusion error rates when comparing determinations according to the *Difficulty of Comparison* rating when the source was present. The percentage of erroneous exclusions in the Difficult to Moderate (>7-14) range was higher as compared to trials that were rated in the Insufficient to Difficult (0-7) and Moderate to Easy (>14-21) range. Furthermore, the exclusion error rate was lowest for latent trials that were rated the easiest to compare.

No erroneous identifications were reported for latent trials that were rated in the Insufficient to Difficult (0-7) range as compared to trials that were rated in the Difficult to Moderate (>7-14) and Moderate to Easy (>14-21) range. Participants reported more inconclusive decisions than correct identifications for latent trials that were rated the most difficult to compare.

Erroneous Identifications of ACE-V Trials

Of the 42 erroneous identifications reported in both Phase 1 and Phase 2, seventeen of these errors occurred during Phase 2 ACE trials. The seventeen erroneous identifications were sent to fourteen of the 63 participants for verification in Phase 3, and fifteen responses for the seventeen erroneous identifications were returned. None of the fourteen participants agreed with the initial erroneous identification; twelve participants disagreed a total of thirteen times and two participants reported an inconclusive decision.

Accuracy of ACE and ACE-V Trials

For ACE examinations, the FPR was 3.0% and FNR was 7.5%. Alternatively, for ACE-V examinations, the FPR was 0.0% and the FNR was 2.9%. The results support the hypotheses that the participants were able to correctly identify or exclude unknown latent print to known standards using both the ACE and ACE-V methodology with a high degree of accuracy.

Significantly Varied Results

To evaluate if the participants reached significantly varied results using the ACE methodology, the latents were grouped by their *Difficulty of Comparison* (source present) and *Strength of Value* (source not present) ratings for comparison purposes. For latents with a *Difficulty of Comparison* rating in the 0-7 range, there was a consensus of inconclusive decisions 76.4% of the time. For latents with a *Difficulty of Comparison* rating in the >7-14 and >14-21 range, there was a consensus of correct identifications 70.2% and 93.0% of the time, respectively. For latents with a *Strength of Value* rating in the 0-7 range, there was a consensus of inconclusive decisions 68.2% of the time and a consensus of correct exclusions in the >7-14 and >14-21 range 64.8% and 83.6% of the time, respectively.

Precision

To determine if the participants would reproduce conclusions from comparisons of unknown latent prints to known standards using the ACE methodology, the results of correct and erroneous identification decisions were evaluated. Participants were able to reproduce a correct identification 94.2% of the time (487 of 517 participant responses) and not reproduce an erroneous identifications 100% of the time (15 of 15 participant responses).

To determine if the participants would repeat their previous conclusions from comparisons of unknown latent prints to known standards using the ACE methodology, the results of correct identifications, erroneous identifications, erroneous exclusions and inconclusive decision were evaluated. Participants repeated their own correct identifications 94.6% of the time (980 of 1,036 participant responses); did not repeat their own erroneous identifications 68.8% of the time (11 of 16 participant responses); and did not repeat their own erroneous exclusions 93.1% of the time (81 of 87 participant responses). Additionally, the participants were almost evenly distributed when presented with their previous inconclusive decision and given the correct source; 49.4% of the participants repeated their own inconclusive decisions (85 of 172 participant responses) and 50.6% of the participants did not repeat their own inconclusive decision (87 of 172 participant responses).

Bias

To determine if the participants would reproduce conclusions from unknown latent prints to known standards using the ACE methodology under biased conditions, the results of correct and erroneous identification decisions were evaluated. Participants were able to reproduce a correct identification 73.0% of the time (178 of 244 participant responses) and to not reproduce an erroneous identifications 96.5% of the time (82 of 85 participant responses).

To determine if the participants would repeat their own conclusions from comparisons of unknown latent prints to known standards using the ACE methodology under biased conditions, the results of correct identifications, erroneous identifications, erroneous exclusions and inconclusive decision were evaluated. Participants repeated their own correct identifications 93.2% of the time (233 of 250 participant responses); repeated their own erroneous identifications 100% of the time (one participant response); and did not repeat their own erroneous exclusions 85.2% of the time (23 of 27 participant responses). Additionally, 60% of the participants repeated

their previous inconclusive decision when given the correct source (33 of 55 participant responses).

B. Implications for Policy and Practice

Sufficiency Determinations

A long standing issue within the latent fingerprint community is that latent print sufficiency determinations are not standardized in terms of a measurable scale. Due to the nature of friction ridge skin, transfer and collection of friction ridge skin impressions and human factors that exist during interpretation, sufficiency determinations should continue to be based on both quantitative and qualitative aspects. A *Strength of Value* rating scale, similar to the one designed for this research, could be utilized by latent print examiners in order to assist them in making appropriate sufficiency determinations.

Accuracy and Precision of ACE and ACE-V

Although this study was not designed to precisely measure how the participants applied the ACE methodology when making their comparisons, data was collected from the participant answer sheets that indicated key components of their Analysis (clarity, anatomical source and certainty of orientation), Comparison (standards used in their comparison), and Evaluation (identification, exclusion, or inconclusive). When taking into account the error rate of erroneous identifications and erroneous exclusions during ACE examinations, the findings support the importance of an independent review of fingerprint conclusions to reduce errors in fingerprint examinations.

Precision of Participant Decisions and the Effects of Contextual Bias

Different trials were sent to participants for verification and second verification reproducibility trials, and the overall results indicate that a contextual bias may have been introduced when participants were presented with two previous conclusions and asked to perform a second verification. Participants who were asked to perform a second verification agreed less often with an initial correct identification and reported more inconclusive decisions. In addition, when participants performing a second verification were presented with an initial erroneous identification, participants reported less inconclusive decisions and were more likely to either agree or disagree with an incorrect identification.

Different trials were also sent to participants to test for repeatability in the form of a verification and second verification. The participants were not made aware that they were verifying or conducting a second verification of their previous conclusions. In testing for repeatability, the effects of contextual bias may have been introduced when participants were presented with two previous conclusions and asked to perform a second verification. Participants performing second verifications repeated their previous erroneous exclusions and inconclusive decisions more often, and were less likely to change these decisions to correct identifications. However, when participants were presented with their initial correct identifications as a second verification, contextual bias did not appear to be a factor as all participants repeated their initial correct identifications at approximately the same rate.

C. Implications for Further Research

The research presented in this study was based on latent prints that were assigned a *Strength of Value* and *Difficulty of Comparison* rating that was created by the researchers. A sufficiency

rating scale that takes into account the level of detail present, amount and type of minutiae, as well as how much weight latent print examiner assign to these features warrants further research to determine what factors are most significant during both the analysis and comparison of latent prints.

Additionally, the empirical data collected in this research study suggests a need for further research into the area of erroneous fingerprint identifications and exclusions under both unbiased and biased conditions. Under unbiased conditions, participants with less latent print examination experience had a higher identification error rate than participants with more experience, but all had a high rate of erroneous exclusions. Results also showed that identification and exclusion error rates were lower when results were independently verified (ACE vs. ACE-V). Under biased conditions, results showed that participants reported more inconclusive decisions when asked to perform a second verification on a correct identification, but more likely to report conclusive decisions when asked to perform a second verification on an erroneous identification. Participants were also more likely to erroneously exclude the correct source when they were given multiple standards to compare against versus when they were asked to make a one to one comparison (single source).

VI. REFERENCES

1. Ashbaugh, D., Quantitative-Qualitative Friction Ridge Analysis. CRC Press. 1999, 156-157.
2. Evett, I.; Williams, R. A Review of the Sixteen Points Fingerprint Standard in England and Wales. In *Proceedings of the International Symposium of Fingerprint Detection and Identifications*; Almog, J., Spinger, E., Eds.; Ne'urim, Israel, 1995, pp 287-304.
3. Gutowski, S. Error Rate in Fingerprint Examination: The View in 2006. *The Forensic Bulletin*, **2006** (Autumn), 18-19.
4. Harold, T., Individualization: Principles and Procedures in Criminalistics. *Lighting Powder Company, Inc.* 1994, 30.
5. Huber, R.A., "The Philosophy of Identification", RCMP Gazette, July-Aug 1972.
6. Langenburg, G. A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability, and Biasability of Conclusions Resulting From the ACE-V Process. *J. For. Ident.*, **2009**, 59 (2).
7. Langenburg, G.; Champod, C.; Wertheim, P. Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology When Conducting Fingerprint Comparisons. *J. For. Sci.*, 2009, Vol. 54, No 3.
8. Latent Prints: A Perspective on the State of the Science”. FBI’s *Forensic Science Communications* (October 2009, Vol. 11, No. 4)
9. “Standards for Minimum Qualifications and Training to Competency for Friction Ridge Examiner Trainees”. Scientific Working Group on Friction Ridge Analysis (SWGFAST, 2010).
10. Wertheim, K.; Langenburg, G.; Moenssens, A. A Report of Latent Print Examiner Accuracy During Comparison Training Exercises. *J. For. Ident.* **2006**, 56 (1), 55-93.
11. Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST), Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint) Document #10, Version 2.0, Accessed 4/1/2014.
12. International Association for Identification (IAI) Report of the Standardization Committee of the IAI – August 1, 1973, Accessed 4/1/2014.
13. Ulery, B.T.; Hicklin, A.R.; Buscaglia, J.; and Roberts, M.A. (2011). Accuracy and Reliability of Forensic Latent Fingerprint Decisions. *Proceedings of the National Academy of Sciences* 108(19): 7733-7738, Appendices 1-26.

14. U.S. Department of Justice, National Institute of Justice, The Fingerprint Sourcebook. *Create Space Independent Publishing Platform*. 2012. Ch. 9, 9.3.1.

VII. DISSEMINATION OF RESEARCH FINDINGS

- The International Forensic Research Institute, 1st and 2nd Annual Forensic Science Symposiums (2012 and 2013, Florida International University, Miami, Florida)
- International Association for Identification 98th Educational Conference (2013, Providence, Rhode Island)
- Florida Division of the International Association for Identification 54th Annual Forensic Training Conference (2013, Miami, FL)
- Two webinars entitled, “Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy, Precision, Reproducibility and Repeatability in Latent Fingerprint Examinations” hosted by RTI International, Center for Forensic Sciences, (2013)

VIII. ACKNOWLEDGMENTS

The authors would like to acknowledge the combined efforts of the consultants on this project; Dr. Glenn Langenburg for his dedication to the improvement of the discipline of fingerprint identification; the volunteers who assisted in creating the tests, as well as the numerous latent fingerprint examiners from across the United States who participated in this study. Without the support and assistance of all of these individuals this project would not have been possible. Their contributions have provided further knowledge and research to assessing the reliability of latent fingerprint examiner decisions.

IX. APPENDIX A – PARTICIPANT EXIT QUESTIONNAIRE

84 Participants completed the Exit Questionnaire

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
The comparisons were representative of actual case work?	0.00%	1.19%	3.57%	57.14%	38.10%
# of responses	0	1	3	48	32
I was given sufficient amount of time to complete the comparisons.	0.00%	2.28%	5.95%	33.33%	58.33%
# of responses	0	2	5	28	49
I made these comparisons with the same care I would with actual case work.	0.00%	4.75%	1.19%	35.71%	58.33%
# of responses	0	4	1	30	49
I am confident in my conclusions.	0.00%	0.00%	1.19%	30.95%	67.86%
# of responses	0	0	1	26	57
I would participate in future research of this kind.	0.00%	1.19%	4.75%	28.57%	65.48%
# of responses	0	1	4	24	55

X. APPENDIX B – SELECT ERRONEOUS IDENTIFICATIONS

Gamma037b – Source C, 8
Latent Score 9.67
Difficulty of Comparison 9.8
94.29% of the time this print was called “Of Value”
Correctly ID: 57 Times
Erroneously Identified: 3
Inconclusive: 11
Erroneously Excluded: 26

Correct Identification



Three latent print examiners erroneously identified this latent to J, #6

Erroneous Identification



Lambda015a – Source Not Present

Latent Score 16

Difficulty of Comparison N/A

100% of the time this print was called “Of Value”

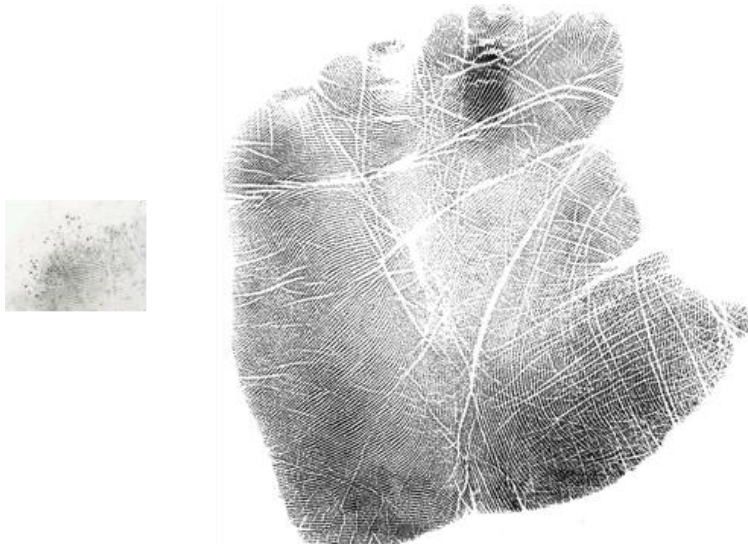
Correctly Excluded: 33

Erroneously Identified: 2

Inconclusive: 2

One latent examiner erroneously identified this latent print to Standard C, left palm. The examiner was uncertain of the orientation.

Erroneous Identification



A second latent examiner erroneously identified this latent print to Standard J, finger 6. The examiner was “certain” of the orientation. The researchers could not verify which orientation the examiner used to effect this erroneous identification.

Erroneous Identification



Gamma039b – Source C, 10

Latent Score 10

Difficulty of Comparison 10.5

88.10% of the time this print was called “Of Value”

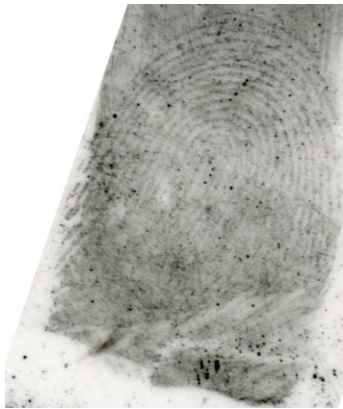
Correctly ID: 32 Times

Erroneously Identified: 3 (2 of the 3 appear to be clerical)

Inconclusive: 2

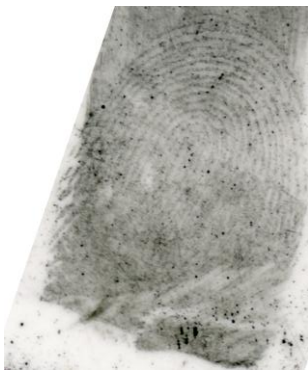
Erroneously Excluded: 1

Correct Identification



One latent print examiners erroneously identified this latent to Standard I, #6

Erroneous Identification



Nu011c – Source Not Present

Latent Score 17

Difficulty of Comparison N/A

97.56% of the time this print was called “Of Value”

Correctly Excluded: 31 Times

Erroneously Identified: 1

Inconclusive: 8

One latent print examiners erroneously identified this latent to Standard J, Right Palm. The examiner was “certain” of the orientation. The researchers could not verify which orientation the examiner used to effect this erroneous identification.

Erroneous Identification

