Article

Assessing Latent Print Proficiency Tests: Lofty Aims, Straightforward Samples, and the Implications of Nonexpert Performance

Brendan Max
Joseph Cavise
Richard E. Gutierrez

Forensic Science Division
Law Office of the Cook County Public Defender
Chicago, IL

Abstract: Proficiency testing has long played a prominent and multifaceted role for the discipline of latent print comparison. Practitioners rely on proficiency testing results when making claims about the overall reliability of the field, as well as when making individual assertions of competency. In fact, in the case of unaccredited laboratories employing uncertified examiners, such testing may ultimately amount to the only objective evidence of competence that judges can consider. Additionally, proficiency testing plays a critical role in laboratory quality assurance programs, where results theoretically can be used to identify gaps in knowledge and training, as well as to supply roadmaps for improvement. But despite these vital functions, there is surprisingly little critical research or analysis by those within the discipline as to whether the current proficiency testing regime is effective and fit for purpose. And what little analysis does exist, mostly from sources external to the latent print comparison community, raises questions about the sufficiency of current tests. In order to more thoroughly investigate the rigor of current latent print proficiency testing, the authors (three assistant public defenders) participated in the Collaborative Testing Services Test No. 18-5161 and completed the test without committing any false positive errors. What follows is a discussion of that participation and its implications. The authors' hope is that the results spark concern about, and reforms to, the current proficiency testing practices in the United States.

Received October 24, 2018; accepted January 23, 2019

Introduction

Discussion about the reliability of latent print evidence (both in general as a discipline and specifically in individual cases) has often involved reference to proficiency testing. With regard to the discipline as a whole, leaders from within the field have used practitioners' near-universal passage of yearly proficiency tests as evidence to assert that the underlying science of latent print comparison is reliable and accurate enough for use in identifying the one true source of a latent print to the exclusion of all other possible sources [1]. Others reference "impressively high" passage rates to support claims about the overall accuracy of examiners engaged in casework [2]. Separate from discipline-wide reliability claims, laboratories rely on proficiency testing as a critical component of quality assurance programs. In theory, proficiency testing results are supposed to provide laboratories with a metric by which they can identify gaps in the knowledge and skill of individual examiners, thereby providing a roadmap for retraining examiners to full competency [3–6].

Perhaps most importantly, at least from the perspective of the legal system, proficiency testing also serves as one (and perhaps the most readily available) means by which judges can fulfill their duty to determine whether a given examiner has sufficiently demonstrated the skill and expertise necessary to reliably (or at least competently) discern the probable source of a pertinent latent print [7, 8]. Generally, judges will permit witnesses to venture beyond merely reporting observations and into the territory of interpreting and opining on the significance of particular facts only when a witness' "specialized knowledge" or skills exceed those of everyday people, allowing them, at least in theory, to accurately formulate opinions that would be beyond the capacities of jurors and thereby "assist the trier of fact" [8–11]. And U.S. courts have long considered the comparison of latent prints an arena in which the assistance of such expert witnesses is not simply useful, but necessary [9, 12, 13], a position recently vindicated by research confirming that fingerprint examiners (at least legitimately competent and experienced ones) do outperform novices and laypeople when comparing the same latent print and known samples [14, 15]. But given the subjectivity presently inherent in the practice of latent print comparison and the heterogeneous performance of examiners during error rate studies, judges cannot simply defer to the overall reliability of the discipline when evaluating individual latent print examiners; they remain in need of effective metrics to appropriately separate the wheat from the chaff [16].

Of concern, however, is the fact that judges engaged in this inquiry have few tools available to help them gauge the competency of latent print examiners. For instance, although the Scientific Working Group on Friction Ridge Analysis, Study, and Technology (SWGFAST), among others, have developed general rubrics for appropriately training latent print examiners [5, 17], they have never concurrently provided a standardized form of competency testing to measure whether examiners have digested such instruction [18], much less a means, vital when evaluating expertise, of measuring whether demonstrated knowledge of various training objectives would actually predict superior performance during casework [19]. And although many from within the forensic science community promote certification (and its attendant testing regime) as one way to satisfy such inquiries [5, 20, 21], the relative dearth of certified latent print examiners [22] means that courts will too often be left with no other option than to look to proficiency testing results when seeking an empirical measure of a proffered expert's performance and abilities.

Yet despite the centrality of proficiency testing to quality assurance and examiner competence, those within the discipline of latent print comparison have engaged in very little critical analysis of current testing practices. SWGFAST has said almost nothing about what standards ought to apply to the content or validation of such exams, making only a few passing comments about appropriate levels of difficulty–for example that "[t]ests should include impressions from different sources that bear resemblances in first and second-level detail"–as well as opining that "[v]alidation is accomplished through predistribution testing of three or more examiners" [3]. And remarkably little (if any) published analysis exists that engages in after-the-fact debriefing of past exams [23, 24]. In fact, the authors could locate only one study designed to answer even the basic question of whether existing proficiency tests so much as discriminate between novices and trained latent print examiners (and thereby serve as appropriate proof of minimal expertise during judicial evaluations of witness qualifications) [25]. Moreover, what commentary can be found (and it largely originates from outside the discipline or outside the United States) has been highly critical of commonly encountered proficiency tests, lambasting them for involving less challenging samples than those routinely encountered during casework, failing to adequately balance target prints with similar distractors, and producing virtually no errors [4, 7, 14, 26–30]. If it is therefore true that, "it is not

clear what these tests say about the proficiency of fingerprint examiners, if they say anything at all" [31], then both latent print examiners and the courts will remain in a shadow of doubt about whether their safeguards for quality are actually buttressed by such tests. Accordingly, in an attempt to shed some light on the sufficiency of the current proficiency testing regime for latent print comparisons in the United States, however modest, and taking seriously the notion that "[u]nderstanding the performance of trained analysts usually requires an understanding of how an untrained individual will perform" [15], the authors decided to use themselves as a control group of sorts and see firsthand whether, despite being attorneys and not latent print examiners, they too could pass a proficiency test. What follows is a summary of the authors' participation in such a test, along with discussion of the authors' observations of the samples the test presented and the inferences warranted by their results.

CTS Latent Print Examination Test No. 18-5161/2/5[1]

The authors are three assistant public defenders who lack the requisite background and training endorsed by discipline guidelines for latent print examiners who are minimally trained to competency. Discipline guidelines describe the recommended minimum training for latent print examiners as including: (1) a bachelor's degree with some "science-related coursework", (2) "thorough instruction" including "mentor sessions, research and papers, presentations, moot courts", (3) successful completion of "written tests and practical exercises", and (4) an apprenticeship for a period of one to two years [17].

The authors' backgrounds do not, in any meaningfully way, adhere to these guidelines. Each has a bachelor's degree in humanities or social science as opposed to a natural or physical science, although because the phrase "science-related coursework" is vague, each arguably met this requirement. As for thorough instruction, none of the authors attended mentorship sessions, and they could not be sure whether they satisfied the requirements of "research and papers" and "presentations" because these terms likewise are vague. None of the authors passed any written tests or practical exercises. And perhaps most

[1] All data regarding the performance of the 353 participants on Latent Print Examination Test No. 18-5161/2/5 is drawn from the Summary Report released by CTS regarding the exam and publicly accessible at https://cts-forensics.com/reports/38161_Web.pdf. Images of latent and known prints from the proficiency test itself have been reproduced and adapted with the permission of Collaborative Testing Services.

importantly, none of the authors participated in an apprenticeship program of any length, and therefore lack the training most directly related to participation in proficiency testing–conducting latent print comparisons with the ACE-V method.

Although clearly failing to meet SWGFAST's minimum standards, the authors did have some training and experience in the discipline. Each author had attended conferences focused on forensic science, which sometimes included discussions of latent print comparison. Each had also reviewed a substantial quotient of research regarding forensic latent print comparison, seeking to identify and familiarize himself with publications in the field for purposes of preparing for motion litigation and cross-examination (relatedly, the authors had all actually litigated multiple cases involving latent print evidence that entailed reviewing the work of examiners, conducting pretrial interviews in order to understand examiner opinions in cases, and sometimes cross-examining them at trial). And finally, each author received approximately four days of training focused specifically on latent print comparison. Most of this training involved discussion of theory and research in the field rather than detailed assessment of actual latent print images. Importantly, none of this training required the authors to engage in a full application of the ACE-V process–assessing a latent print for value, identifying reliable comparison features, and searching across multiple exemplars in an attempt to locate a possible association. Despite this exposure to latent print concepts, the authors would hope that no one inside or outside of the forensic latent print community would accept them as competent experts in the field. In fact, the first time that the authors ever sought to conduct a complete latent print examination was during their participation in the Collaborative Testing Services (CTS) Latent Print Examination Test No. 18-5161/2/5.

The authors selected a CTS test[2] for their experiment because that company has emerged at the forefront of designing and disseminating proficiency tests to forensic practitioners and labs throughout the United States [11]. Each author paid for and received a separate test packet, with an individual participation

[2] Forensic laboratories have limited choices in the marketplace for fingerprint proficiency test providers, and even fewer if a laboratory seeks a proficiency test provider with ISO 17043 accreditation. With most laboratories turning to CTS for fingerprint proficiency test services, an assessment of CTS proficiency testing should be of paramount concern to the discipline.

code for the test[3]. Each test packet consisted of photographs of 12 latent prints and 4 sets of photographs representing print exemplars for comparison. Upon receipt of the test packet, each author completed the test independently. This was done over the course of approximately one month, with each author working this task into otherwise busy litigation schedules. This resulted in the authors conducting their analyses at different times and places from each other. Finally, each author independently completed an answer sheet provided by CTS and independently reported his answers through the CTS on-line portal. To assist in the examination process, the authors purchased two plastic hand-held magnifiers.

The authors each reached 11 correct (consensus) results out of 12 comparisons. Importantly, none of the authors reported a response that resulted in a false positive error (meaning for the combined 36 opportunities on the test to identify an incorrect source of the 12 latent prints, the authors never did so). Each author did, however, reach one nonconsensus response, answering "not identified" to a latent print whose mated pair was included among the test's exemplars. But that comparison (involving the latent print designated Q11) also stumped the highest number of participants generally (eliciting a nonconsensus response from 16 of the 353 participants).

Most participants found the CTS test no more challenging than the authors did. A full 91% of participants (320 of 353) completed the CTS test without committing any errors. That percentage rises to 96% when including participants who (like the authors) made no more than one error (340/353). Moreover, the test was clearly not designed to generate many false positive errors. Overall, the 353 participants provided false positive responses (matching the latent to the wrong source altogether or matching it to the wrong finger or palm of the correct source) in just 19 of 4236 comparisons (or just 0.45% of the time). And the majority of the false positive responses (12 of the 19) were attributable to 3 outlier participants who reported 3, 4, and 5 false positive responses, respectively, out of 12 comparisons each. Excluding these outliers (who one would hope are not practicing latent print examiners), the remaining 350 participants provided only 7 false positive responses in 4200 comparisons (or just 0.17% of the time).

[3] The authors' proficiency test results can be accessed under Web Codes AYRKMK, F7VF9F, and ZHFPEU, at https://cts-forensics.com/reports/38161_Web.pdf.

Discussion

Initially, the authors' results (albeit preliminary and not reproduced or repeated across multiple studies or CTS exams) suggest that the current proficiency testing regime for latent print examiners does not adequately discriminate between experts and novices. It should not, in other words, have been possible for three attorneys with little training in latent print comparison to independently complete a proficiency test had it been appropriately designed to vet the competence of trained forensic professionals. And it bears mentioning that researchers reached similar conclusions after simulating novice performance on one of CTS's 2016 latent print proficiency tests, ultimately explaining that "...discrimination between novices and experts is not possible at an individual level. That is, if someone received a perfect score, it is not clear whether they are an expert or a novice, since both novices and examiners mostly answered 12 questions correctly." [25]

Moreover, given the paucity of false positive errors across multiple years of CTS tests [32], the simple fact of the authors' success on Test No. 18-5161, even standing alone, should raise troubling questions about the adequacy of that company's testing instruments for latent print examiners beyond just its 2018 edition. But understanding specifically why and how the authors (along with the vast majority of other test takers) were able to avoid committing any false positive errors nonetheless warrants a deeper look at the samples comprising the test itself and an analysis of the challenges they failed to pose. And such closer inspection ultimately reveals one compelling explanation for the authors' results: as CTS and its representatives have already suggested would be the case by noting the commercial pressure to produce easy tests [7], the samples included in Test No. 18-5161 were "more artifact than real world" [33], including few, if any, of the aspects of latent print comparison that challenge examiners and produce errors during casework.

The task of evaluating the difficulty of specific latent print comparisons is complicated by a lack of objective metrics as well as a limited understanding of "how the visual characteristics of fingerprint pairs relate to comparison difficulty" [34]. Nevertheless, latent print examiners have identified some characteristics that increase the difficulty of comparison work, require extensive training to overcome, and can produce errors when examiners lack such training [16]. For purposes of this article, the authors focus on three such challenges involving latent prints where (1) a large number of Level 2 features are not available for comparison, (2) what Level 2 features do exist

are rendered ambiguous and less reliable by quality issues, and (3) available Level 2 features appear in highly similar configurations compared to two or more potential candidates (close nonmatches) [5, 35–38]. Participants in Test No. 18-5161 were forced to confront none of these sources of complexity.

To begin with, Test No. 18-5161 contained no latent prints with a borderline number of comparison features. Although the quantity of features in a latent print is not the sole indicator of the complexity of the examination, it is both the most important one and the only objective one (the subjective assessment of "quality" being the other recognized indicator) [36]. And although latent print examiners in the United States do not officially follow a point standard–a number of identifiable features below which an examiner will refuse to offer an association opinion–research indicates a de facto threshold of about eight features [35]. For the 10 comparisons on the CTS test in which they reached association opinions, the authors identified an abundance of features on each latent print (Table 1). In fact, the one author who consistently documented the most features for comparison identified an average of 19 Level 2 features across the 10 comparisons in which he reached association opinions. And at times, the number of available features soared well beyond the threshold generally needed by latent print examiners making associations, with the authors identifying 30, 23, and 24 features, respectively, in Latent Print Q10. But even in the sample (Latent Print Q1) containing the fewest number of Level 2 details, the authors still identified 14, 12, and 14 features, respectively. In other words, none of the comparisons comprising Test No. 18-5161 put the authors (or, one hopes, any other test participants) in the difficult position of trying to decide whether similar constellations of only seven or eight features represented a true association.

	Author 1 # of Features Identified	Author 2 # of Features Identified	Author 3 # of Features Identified
Latent Q1	14	12	14
Latent Q2	14	12	13
Latent Q4	19	13	24
Latent Q5	14	21	22
Latent Q6	15	14	15
Latent Q7	21	18	26
Latent Q8	16	13	18
Latent Q9	13	13	20
Latent Q10	30	23	24
LatentQ12	15	12	17

Table 1

Quantity of identifiable Level 2 features documented by each author.

The diverse array of factors affecting quality, another recognized indicator of latent print comparison difficulty [16, 36], complicates any concise discussion of the issue, but it is nevertheless clear to the authors that Test No. 18-5161 did not confront participants with serious challenges in that regard. For example, no latent print examiner would likely view Latent Print Q10 (Figure 1) as anything other than high quality given the abundance of distinct Level 2 detail and the readily identifiable presence even of creases across the finger. But more than simply containing a few crystal clear latent prints, Test No. 18-5161 also failed to challenge test takers by including samples displaying well-recognized issues of clarity that have been shown to trouble examiners during casework and error rate studies. One such type of comparison would have involved latent prints that exhibit tonal reversal. The phenomenon of tonal reversal sometimes causes latent print examiners to mistake furrows for ridges, thereby leading the comparison process astray [38]. The particular challenge with tonally reversed prints was well documented in the FBI Black Box Study, where 5 prints that were tonally reversed (out of a pool of 744 latent prints) generated 5 of the 6 false positive results across 4083 comparisons [16]. But, as with the other types of challenging prints mentioned above, Test No. 18-5161 did not present this challenge for participants. Taken together with the quantity of features in the latent samples from Test No. 18-5161, the lack of quality issues (including tonal reversal) means that exam participants were asked to work only with prints that SWGFAST would define as "non-complex" [36]. Participants were not, in other words, asked to perform the types of comparisons that, during casework, might actually spark disagreement, provoke error, or justify additional safeguards (i.e., enhanced documentation or verification) [36].

Finally, separate from latent prints with few or ambiguous comparison features, the discipline has recognized that cases involving close nonmatches likewise represent a significant challenge to examiners [5, 37]. A close nonmatch is an exemplar print that is not the true source of the latent print, but that nonetheless shares a significant amount of comparison features in a constellation similar to the latent print. Because of this similarity, comparisons involving close nonmatches can be difficult and can fool examiners into false positive errors. In fact, it was the existence of a close nonmatch print in the Madrid bombing case that fooled multiple experienced latent print examiners into misidentifying Brandon Mayfield as the source and generated

the most infamous false positive in the discipline [37]. Past CTS proficiency tests have included such challenges–for example its 1995 test relied on prints from monozygotic twins to vet examiners' ability to avoid false positives when faced with highly similar distractor prints [5, 23]. But on Test No. 18-5161, the authors did not encounter any comparisons that involved close nonmatches. The fact that the untrained authors did not commit a single false positive error in 36 combined comparisons supports the conclusion that none of the incorrect distractor choices were similar enough to the true source to cause confusion. Although only CTS can describe the steps it took to replicate the close nonmatch scenario from casework, the authors did not encounter distractor prints that shared a significant number of comparison features (8 or more) with any of the 11 latent prints for which the true source was present. Rather, the closest distractor was often very easy to distinguish and did not represent a close nonmatch.

Using Latent Print Q10 from Figure 1 as an example, none of the distractors should have fooled any participants, and none can be considered a true close nonmatch (Figure 2). All but two of the participants correctly identified Item D, Right Index as the source of Q10 (one participant incorrectly reported that the source of Q10 was not present in the four exemplars provided). Although 1 of the 353 participants erroneously selected the right index finger from Item A as the source of Q10, it is hard to see how this could have happened. These prints look nothing like each other and do not even share similar Level 1 appearance. Aside from this aberrant result, the closest distractor for Q10 was the left little finger from Item B. But even acknowledging the similar Level I appearances, the differences between comparison features in Q10 and the little finger from Item B are many and obvious. Although there is some subjectivity in determining whether a latent print is similar enough to the true source to be considered a "close nonmatch", this distractor should not have and ultimately did not fool any of the 353 participants.

Ultimately, the burden, scientifically speaking, falls on the designer of proficiency tests, as well as the proponents of any particular inference drawn from such instruments (i.e., any examiner or attorney who would deploy passing marks from CTS testing as proof of competency or robust quality assurance), to demonstrate the validity of said exams [39]. But the authors' performance on Test No. 18-5161–especially when coupled with the historically small number of false positive errors on other CTS exams and the fact that its 2018 test failed to offer the types of challenging examinations encountered in casework–casts

Figure 1
Latent image for Q10.

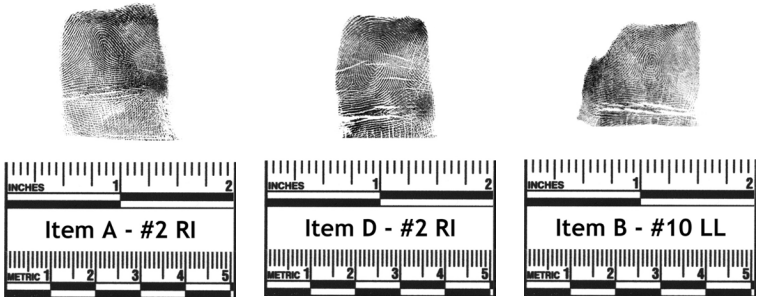

Figure 2
Item A, right index. *True source of Q10 latent.* *Item B, left little.*

doubt on whether any such inferences are actually warranted. With regard to quality assurance, labs that employ CTS testing as an important component of their QA programs are operating under a system that sets that bar too low and that is not sensitive enough to identify real breaking points in casework. These labs should not claim to have robust QA systems that can identify and remediate gaps in training [3–6]. And with regard to competency, judges and attorneys in the criminal justice system must realize that latent print examiners (especially uncertified ones) who assert competency in large part based on successful CTS testing are doing so based on completion of simplistic comparison tasks that say little to nothing about how examiners might perform under more challenging casework conditions [4]. It therefore appears that CTS tests (at minimum Test No. 18-5161) offer little value to forensic laboratories or the consumers of their work.

Conclusions

When evaluating purported expertise, social scientists–who, much like the law, emphasize the notion that experts should be capable of superior performance on domain tasks (i.e., above and beyond that of laypeople and novices when confronted with highly similar but ultimately distinct cases), strongly suggest relying on empirical measures and testing as opposed to credentials and experience, because the latter categories are at best "uncertain predictor[s] of degree of expertise" [19, 40–42]. And commentators (as well as, at times, the courts) have begun to come around to this view and accordingly emphasize review of proficiency testing results [11, 43, 44]. In fact, far from rejecting empirical measures of competence, even the forensic science community has widely embraced such approaches (including proficiency testing), considering them "hierarchically greater" in value than the other potential metrics just mentioned [3, 5, 20, 45–47]. But the authors' results should spark concern that the proficiency tests administered by CTS, and so widely relied on by the latent print comparison community, are not equal to the vital tasks carved out for them and do not warrant the type of esteem surely bestowed on them by judges and juries when they hear merely that examiners have taken and passed such exams every year. It is therefore well past time for the latent print comparison community to turn greater attention and resources to the task of demanding and validating a rigorous proficiency testing regime.

Other forensic communities have already taken vital steps in that vein. For instance, the forensic DNA community participated in the NIST Mix2013 study, which contained challenging mixture interpretations. Large percentages of participants provided erroneous responses to these mixture interpretations, which in turn prompted robust discussions in the community about subjectivity and the true outer limits of reliable DNA comparisons [48]. European firearms examiners have similarly produced a vigorous and robust proficiency testing alternative that has highlighted gaps in examiner understanding of subclass characteristics and enabled retraining of examiners to prevent casework errors [49]. In fact, the latent print community can simply look inward to older CTS tests from the 1990s. The CTS latent print test administered in 1995, for example, was designed to replicate challenging casework; it included latent prints made in blood, prints exhibiting tonal reversal, and close nonmatch distractors obtained from the twin brother of the true source of the latent print [5, 23]. The results were nothing like the universal passage on the 2018 version, with only 44% of the 156 participants providing correct responses to the seven examinations presented. For one challenging comparison involving a close nonmatch twin distractor, 29 of the 156 participants (18%) erred. Importantly, these results spurred on discussion and action in the latent print community, including movements to develop national guidelines for training and quality assurance [23].

Even beyond these practical examples, the latent print comparison community can look to the field of psychometrics for decades worth of literature specifically grappling with the vital question of how to evaluate instruments that themselves purport to assess competent performance on particular tasks [39]. In contrast to the fairly limited inquiry into validity proposed by SWGFAST, that literature suggests a multifaceted, rigorous, and ongoing approach to evaluating the inferences drawn from any particular test [39, 50, 51]. Whereas other scientific fields have relied heavily on psychometric insights when developing their own competency exams [52] and the NIST Working Group on Human Factors years ago suggested that latent print examiners follow suit [5], CTS has seemingly never done so [26]. It is time for such insularity to change.

Should the latent print community decide to undertake a meaningful review of current proficiency testing practices, they would be following in the footsteps of other scientific communities who undertook this important process. For example, the

clinical testing laboratory community, engaged in mandatory proficiency testing since 1988, undertook a significant review and reform process in 2008 [53]. The clinical community sought answers to important questions about their proficiency testing system, such as: (1) what are the attributes of a good proficiency test program, (2) what is the evidence that the current approach is effective, (3) are existing tests rigorous enough, (4) are tests calibrated to identify subpar laboratory practices and identify needed reforms, and (5) do test results provide a basis for the effective education of practitioners? This inquiry required a substantial commitment to scientific advancement and resulted in 21 detailed recommendations for improving then-existing proficiency test practices. Until the forensic latent print community can supply meaningful answers to similar questions about CTS proficiency testing, participation in these tests should be considered little more than window dressing.

For further information, please contact:

Brendan Max
Forensic Science Division
Cook County Public Defender Office
69 West Washington, Floor 17
Chicago, IL 60602
brendan.max@cookcountyil.gov

References

1. Reznicek, M.; Ruth, R. M.; Schilens, D. M. ACE-V and the Scientific Method. *J. For. Ident*. **2010**, *60* (1), 87–103.
2. Stevenage, S. V.; Pitfield, C. Fact or Fiction: Examination of the Transparency, Reliability and Sufficiency of the ACE-V Method of Fingerprint Analysis. *For. Sci. Int*. **2016**, *267*, 145–156.
3. SWGFAST. Document #13: Standard for Friction Ridge Comparison Proficiency Testing Program (Latent/Tenprint). November 2012. http://clpex.com/swgfast/documents/proficiency/121124_Standard-Proficiency-Testing_2.0.pdf (accessed October 2018).
4. Thompson, W.; Black J.; Jain, A.; Kadane, J. *Forensic Science Assessments: A Quality & Gap Analysis: Latent Fingerprint Examination*. American Association for the Advancement of Science: Washington, D.C., 2017.

5. National Institute of Standards and Technology (NIST), Expert Working Group on Human Factors in Latent Print Analysis. *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*; NISTIR #7842; U.S. Department of Commerce: Gaithersburg, MD, 2012.
6. Gray, M. L. Quality Assurance. In *The Fingerprint Sourcebook*; McRoberts, A., McRoberts, D., Eds.; National Institute of Justice: Washington, D.C., 2011; Chapter 12.
7. President's Council of Advisors on Science & Technology. *Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*; Executive Office of the President of the United States: Washington, D.C., 2016.
8. Brown, H. Eight Gates for Expert Witnesses. *Houston Law Review* **1999**, *36* (2), 743–882.
9. Lawson, T. F. Can Fingerprints Lie?: Re-Weighing Fingerprint Evidence in Criminal Jury Trials. *Am. J. Criminal Law* **2003**, *31* (1), 1–66.
10. *Rule 702. Federal Rule of Evidence*; House of Representatives - The Committee on the Judiciary, U.S. Government Printing Office: Washington, DC, 2011.
11. Illinois Rule of Evidence 702. Adopted September 27, 2010, eff.January 1, 2011.
12. Illinois v. Palmer, Illinois Appellate Court: 181 Ill. App. 3d 504 (1989).
13. United States v. Reaux, United States District Court for the Eastern District of Louisiana: No. 01-071 SECTION: "R" (2); 2001 U.S. Dist. LEXIS 11883 (July 31, 2001).
14. Thompson, M. B.; Tangen, J. M.; McCarthy, D. J. Expertise in Fingerprint Identification. *J. For. Sci*. **2013**, *58* (6), 1519–1530.
15. Langenburg, G. A Critical Analysis and Study of the ACE-V Process. Doctoral Dissertation, University of Lausanne, Switzerland, 2012.
16. Ulery, B. T.; Hicklin, R. A.; Buscaglia, J.; Roberts, M. A. Accuracy and Reliability of Forensic Latent Fingerprint Decisions. *Proc. Nat. Acad. Sci.* **2011**, *108* (19), 7733–7738.
17. SWGFAST. Document #18: Standards for Minimum Qualifications and Training to Competency for Friction Ridge Examiner Trainees (Latent/Tenprint). November 2012. http://clpex.com/swgfast/documents/qualifications-competency/121124_Qualifications-Training-Competency_2.0.pdf (accessed October 2018).

18. SWGFAST. Document #102: Recommendations for Competency Testing of Noncertified Latent Print Examiners (Latent) Position Statement. November 2012. http://clpex.com/swgfast/Comments-Positions/121124_Competency-Testing_4.0.pdf (accessed October 2018).
19. Shanteau, J.; Weiss, D. J.; Thomas, R. P.; Pounds, J. How Can You Tell if Someone is an Expert? Empirical Assessment of Expertise. In *Emerging Perspectives on Judgment and Decision Research*; Schneider, S. L.; Shanteau, J., Eds.; Cambridge University Press: Cambridge, U.K., 2003; pp 620–640.
20. SWGFAST. Position Statement on Strengthening Forensic Science in the United States: A Path Forward. August 2009. http://clpex.com/swgfast/Comments-Positions/SWGFAST_NAS_Position.pdf (accessed October 2018).
21. National Commission on Forensic Science. Views of the Commission: Certification of Forensic Science Practitioners. December 2016. https://www.justice.gov/archives/ncfs/page/file/888671/download (accessed October 2018).
22. United States v. Herrera, United States Court of Appeals for the Seventh Judicial Circuit: 704 F.3d 480 (2013).
23. Grieve, D. Possession of Truth. *J. For. Ident*. **1996**, *46* (5), 521–528.
24. Haber, L.; Haber, R. N. Error Rates for Human Latent Fingerprint Examiners. In *Automatic Fingerprint Recognition Systems*. Ratha, N.; Bolle, R., Eds.; Springer: New York, 2004; pp 345–348.
25. Luby, A. S.; Kadane, J. B. Proficiency Testing of Fingerprint Examiners with Bayesian Item Response Theory. *Law, Probability, & Risk* **2018**, *17* (2), 111–121.
26. Haber, L; Haber, R. N. Scientific Validation of Fingerprint Evidence under Daubert. *Law, Probability and Risk* **2008**, *7* (2), 87–109.
27. Koehler, J. J. Forensics or Fauxrensics? Ascertaining Accuracy in the Forensic Sciences. *Ariz. St. L. J.* **2017**, *49* (4), 1369–1416.
28. Saks, M. J.; Koehler, J. J. The Coming Paradigm Shift in Forensic Identification Science. *Sci*. **2005**, *309* (5736), 892–895.
29. Schwartz, A. Challenging Firearms & Toolmark Identification–Part II. *The Champion,* **2008**, *32* (9), 44–52.
30. Kennedy, D.; Merrill, R. A. Assessing Forensic Science. *Issues in Sci. & Tech*. **2003**, *20* (1).
31. Tangen, J. M.; Thompson, M. B.; McCarthy, D. J. Identifying Fingerprint Expertise. *Psych. Sci*. **2011**, *22* (8), 995–997.

32. Garrett, B. L.; Mitchell, G. The Proficiency of Experts. Univ. of Penn. Law Review, July 2018, Vol. 66, pp 901–960.
33. Collaborative Testing Services, Inc. Statement on the Use of Proficiency Testing Data for Error Rate Determinations, March 2010, pp 1–3.
34. Mnookin, J.; Kellman, P. J.; Dror, I.; Erlikhman, G.; Garrigan, P.; Ghose, T.; Metler, E.; Charlton, D. Error Rates for Latent Fingerprinting as a Function of Visual Complexity and Cognitive Difficulty. National Institute of Justice Award 2009-DN-BX-K225. 2016.
35. Langenburg, G.; Champod, C.; Thibault, G. Informing the Judgments of Fingerprint Analysts Using Quality Metric and Statistical Assessment Tools. *For. Sci. Int*. **2012**, *219* (1–3), 183–198.
36. SWGFAST. Document #10: Standards for Examining Friction Ridge Impressions and Resulting Conclusions. March 2013. http://clpex.com/swgfast/documents/examinations-conclusions/130427_Examinations-Conclusions_2.1.pdf (accessed October 2018).
37. Office of the Inspector General Oversight and Review Division. A Review of the FBI's Handling of the Brandon Mayfield Case; Office of the Inspector General, Oversight and Review Division, U.S. Department of Justice: U.S. Government Printing Office: Washington, D.C., 2006; 1–330.
38. Castellon, S. Clues in Friction Ridge Comparisons: Tonal Reversals. *J. For. Ident*. **2014**, *64* (3), 223–237.
39. Sireci, S. G. On Validity Theory and Test Validation. *Educational Researcher* **2007**, *36* (8), 477–481.
40. Weiss, D. J.; Shanteau, J. Decloaking the Privileged Expert. *J. Management & Organization* **2012**, 18 (3), 300–310.
41. Ericsson, K. A.; Charness, N. Expert Performance: Its Structure & Acquisition. *American Psychologist* **1994**, *49* (8), 725–747.
42. Weiss, D. J.; Shanteau, J. Empirical Assessment of Expertise. *Human Factors* **2003**, *45* (1), 104–116.
43. Federal Judicial Center. *Reference Manual on Scientific Evidence*; National Academies Press: Washington, D.C., 1994.
44. United States v. Mitchell, United States Court of Appeals for the Third Judicial Circuit: 365 F.3d 215 (2004).

45. Organization of Scientific Area Committees: Friction Ridge Subcommittee. Response to Call for Additional Reference Regarding: President's Council of Advisors on Science and Technology. Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. December 2016. https://theiai.org/docs/2.20161214_PSAC-FR_PCAST_response.pdf (accessed October 2018).
46. National Commission on Forensic Science. Views of the Commission: Proficiency Testing in Forensic Science. 2016. https://www.ascld.org/wp-content/uploads/2016/03/Adjudication-Summary-of-Proficiency-Testing-in-Forensic-Science-Views-Do....pdf (accessed October 2018).
47. American Society of Crime Laboratory Directors. Board Comments on National Commission on Forensic Science. Views of the Commission: Proficiency Testing. December 2015. https://www.ascld.org/wp-content/uploads/2016/01/Proficiency-Testing-in-Forensic-Science.pdf (accessed October 2018).
48. Coble, M. D.; Butler, J. M. January 2015. DNA Mixture Interpretation: State-of-the-Art. https://strbase.nist.gov/pub_pres/ASCLD-LAB-Jan2015-CobleButler.pdf.
49. Pauw-Vugts, P.; Walters, A.; Øren, L.; Pfoser, L. FAID2009: Proficiency Test and Workshop. *A.F.T.E. Journal* **2013**, *45* (2), 115–127.
50. Messick, S. Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher* **1989**, *18* (2), 5–11.
51. Gyll, S.; Ragland, S. Improving the Validity of Objective Assessment in Higher Education: Steps for Building a Best-in-Class Competency-Based Assessment Program. *J. Competency-Based Ed.* **2018**, *3* (1), 1–8.
52 Taib, F.; Yusoff, M. S. B. Difficulty Index, Discrimination Index, Sensitivity and Specificity of Long Case and Multiple Choice Questions to Predict Medical Students' Examination Performance. *J. Taibah Univ. Med. Sci*. **2014**, *9* (2), 110–114.
53 Peterson, J. C.; Hill, R. H.; Black, R. S.; Winkelman, J.; Tholen, D. Identifying Best Practices in Laboratory Medicine: Review of Proficiency Testing Services for Clinical Laboratories in the United States–Final Report of a Technical Working Group. April 2008. https://www.cdc.gov/labbestpractices/pdfs/Proficiency-Testing-report-2007.pdf (accessed October 2018).