

9-24-2020

Implementing blind proficiency testing in forensic laboratories: Motivation, obstacles, and recommendations

Robin Meija
Carnegie Mellon University

Maria Cuellar
University of Pennsylvania

Jeff Salyards
Compass Scientific Consulting

Follow this and additional works at: <https://lib.dr.iastate.edu/csafepubs>



Part of the [Forensic Science and Technology Commons](#)

Recommended Citation

Meija, Robin; Cuellar, Maria; and Salyards, Jeff, "Implementing blind proficiency testing in forensic laboratories: Motivation, obstacles, and recommendations" (2020). *CSAFE Publications*. 74.
<https://lib.dr.iastate.edu/csafepubs/74>

This Article is brought to you for free and open access by the Center for Statistics and Applications in Forensic Evidence at Iowa State University Digital Repository. It has been accepted for inclusion in CSAFE Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Implementing blind proficiency testing in forensic laboratories: Motivation, obstacles, and recommendations

Abstract

Regular proficiency testing of forensic examiners is required at accredited laboratories and widely accepted as an important component of a functioning quality assurance program. Yet, unlike in other testing industries, the majority of forensic laboratories testing programs rely entirely on declared proficiency tests. Some laboratories, primarily federal forensic facilities, have adopted blind proficiency tests, which are also used in the medical and drug testing industries. Blind tests offer advantages. They must resemble actual cases, can test the entire laboratory pipeline, avoid changes in behavior from an examiner knowing they are being tested, and are one of the only methods that can detect misconduct. However, the forensic context present both logistical and cultural obstacles to the implementation of blind proficiency tests. In November 2018, we convened a meeting of directors and quality assurance managers of local and state laboratories to discuss obstacles to the adoption of blind testing and assess successful and potential strategies to overcome them. Here, we compare the situation in forensic science to other testing disciplines, identifying obstacles to the implementation of blind proficiency testing in forensic contexts, and proposing ways to address those issues and increase the ecological validity of proficiency tests at forensic laboratories.

Keywords

Proficiency test, Blind proficiency testing, Error; Mistake; Misconduct

Disciplines

Forensic Science and Technology

Comments

This article is published as Mejia, Robin, Maria Cuellar, and Jeff Salyards. "Implementing blind proficiency testing in forensic laboratories: Motivation, obstacles, and recommendations." *Forensic Science International: Synergy* 2 (2020): 293-298. Posted with permission of CSAFE.

Implementing blind proficiency testing in forensic laboratories: Motivation, obstacles, and recommendations

Robin Mejia ^{a,*}, Maria Cuellar ^b, Jeff Salyards ^c

^a Carnegie Mellon University Department of Statistics and Data Science, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA

^b University of Pennsylvania Department of Criminology, USA

^c Compass Scientific Consulting, LLC, USA

ARTICLE INFO

Article history:

Received 3 September 2020

Accepted 3 September 2020

Available online 24 September 2020

Keywords:

Proficiency test

Blind proficiency testing

Error

Mistake

Misconduct

ABSTRACT

Regular proficiency testing of forensic examiners is required at accredited laboratories and widely accepted as an important component of a functioning quality assurance program. Yet, unlike in other testing industries, the majority of forensic laboratories testing programs rely entirely on declared proficiency tests. Some laboratories, primarily federal forensic facilities, have adopted blind proficiency tests, which are also used in the medical and drug testing industries. Blind tests offer advantages. They must resemble actual cases, can test the entire laboratory pipeline, avoid changes in behavior from an examiner knowing they are being tested, and are one of the only methods that can detect misconduct. However, the forensic context present both logistical and cultural obstacles to the implementation of blind proficiency tests. In November 2018, we convened a meeting of directors and quality assurance managers of local and state laboratories to discuss obstacles to the adoption of blind testing and assess successful and potential strategies to overcome them. Here, we compare the situation in forensic science to other testing disciplines, identifying obstacles to the implementation of blind proficiency testing in forensic contexts, and proposing ways to address those issues and increase the ecological validity of proficiency tests at forensic laboratories.

© 2020 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Proficiency tests are a core component of quality assurance at testing laboratories [1]. At drug testing laboratories, samples with known levels of alcohol or drugs are provided to examiners for analyses in order to determine whether examiners produce an accurate result [2]. Similar testing occurs in medical laboratories [3]. Such testing is important because in the context of daily work, examiners produce results identifying what is in a submitted sample, which is not known in advance. Proficiency tests enable the evaluation of the performance of an examiner, or a laboratory pipeline, with a sample where the ground truth is known.

Testing program requirements vary by between disciplines such as drug, medical, and forensic testing, following discipline-specific norms as well as legal and accreditation guidelines. In declared (or open) proficiency tests, tests are provided labeled as tests, and often address a specific component of an analysis rather than the

entire process. Blind proficiency tests involve samples that are submitted through the normal analysis pipeline as if they were real cases, requests, or tenders. In blind tests, the examiners conduct the analysis under the assumption they are working on real samples [4]. Only after the work is completed do they learn that a case was a proficiency test.

The goal of forensic science service providers is to produce accurate and reliable results. To fully appreciate the benefits and limitations of proficiency tests it is useful to consider the types of outcomes (i.e., conforming vs nonconforming work) and the types of inaccurate results (i.e., type I and II errors, mistakes, malpractice, and misconduct) that can occur in analyses (see Fig. 1). Conforming work means the examiner properly followed the examination method without any deviations. Nonconforming work refers to examinations where the examiner deviated from the method. Deviations that result from innocent clerical errors (e.g., transposition of numbers) are labeled as mistakes. Mistakes are easily detectable, even by the examiner who made the mistake. Deviations that are the result of poor or incomplete training are best referred to as malpractice. Malpractice can be detected by a robust technical review from a more qualified examiner. Deviations that are deliberate

* Corresponding author.

E-mail address: rmejia@andrew.cmu.edu (R. Mejia).

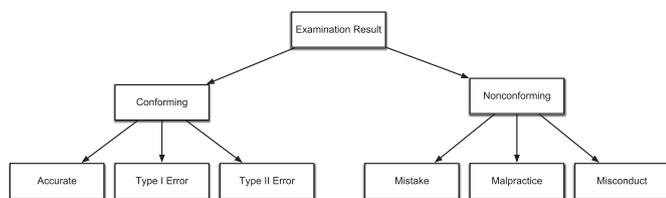


Fig. 1. Possible results of a forensic analysis.

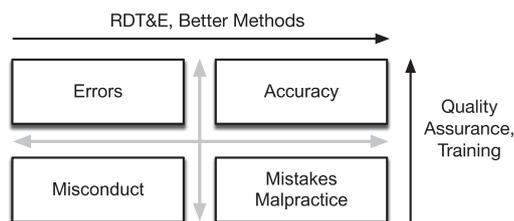


Fig. 2. Role of quality assurance and methods development in improving results.

are described as misconduct. Misconduct can be very difficult to detect because the offending examiner may take steps to make the work appear to be conforming. Fig. 1 depicts these categories. It also illuminates the important point that conforming work does not always lead to accurate results. Forensic examiners might perform an inferior method very well (i.e. with no mistakes or malpractice) but still produce inaccurate results. Proper method validation studies can identify the limitations and accuracy of the method. (These limitations are often expressed as false-positive and false-negative rates, Receiver Operating Characteristics (ROC) curves, or Detection Error Tradeoff (DET) curves.)

In the effort to maximize accuracy, the forensic science community can engage in a few lines of effort. First, laboratories can implement the best methods available (see Fig. 2). Second, enhanced research and development can lead to the creation of better methods with lower error rates and higher accuracy. Third, robust quality assurance programs can detect mistakes. And fourth, training can stave off malpractice. Misconduct, however, is often neither detected nor improved by these efforts.

Proficiency testing plays an important role in promoting accuracy [5]. Test results indicate whether the result of an analyses is correct or not, and an examination of the full test process can provide information on whether the process was conforming or non-conforming, and what steps are needed to improve. Furthermore, blind proficiency testing is one of the few strategies that can reveal instances of misconduct.

Declared tests can provide important checks on examiner’s results and processes. Using tests from a third-party vendor ensures objectivity in scoring and enables inter-lab analyses, and, importantly, meets accreditation requirements as specified in ISO17025.¹ At a meeting discussed later in this paper, quality assurance managers from a range of forensic laboratories indicated that the majority of labs rely on tests purchased from third-party vendors, such as CTS or Forensic Advantage [6].

However, for some disciplines, commercial forensic proficiency tests have been shown to differ substantially from casework, both in terms of tasks and difficulty [7]. Also, commercial tests may only

target a part of the forensic analysis pipeline. For example, a proficiency test for a latent print examiner may provide two or more prints to compare and ask the examiner to determine whether they are of sufficient quality to render a decision and what that decision would be. Those are important steps in the latent print analysis pipeline, but they do not cover all of the work an examiner does on a real case or all the areas where errors could occur between evidence submission and the release of a report. Additionally, recent studies have indicated that latent prints used in proficiency tests by CTS in recent years are of higher quality than those seen in casework [8], and that latent print proficiency exams are now less challenging than they used to be [9]. This means that the tests used to monitor lab quality do not necessarily assess how latent print examiners perform the types of comparisons they see in real casework.

In summary, the tests used to monitor lab quality may not be assessing forensic analyses as they are applied in the field. Further, research has found that examiners behave differently during proficiency testing than during routine analyses, e.g., dedicating additional time to analyses [10]. A recent survey of latent print examiners confirms this behavior [11].

Studies from other testing industries have shown that both behavior and results can differ when examiners are given declared vs. blind proficiency tests. Two studies in drug testing labs in the 1970s compared blind and declared proficiency tests at 24 and 10 laboratories, respectively, and found that false negatives were higher in the blind tests compared to when laboratories knew they were being tested. That is, in cases where examiners did not know they were being tested, they missed more drug samples. In the first study, false positives (incorrectly finding a drug where none was present) were also higher, though in the second study, false positives went down [2]. A 2001 study comparing blind and declared proficiency tests in blood lead testing programs at two large state laboratories found error rates were higher in the blind tests and suggested that laboratories were making special efforts when analyzing known proficiency test samples [12]. Today, the Mandatory Guidelines for Federal Workplace Drug Testing Programs require participating laboratories to conduct blind testing.

However, blind proficiency testing is the exception at forensic laboratories. In 2014, 98% of forensic labs reported conducting some kind of proficiency test, but only 10% conducted blind proficiency tests, the same percentage that had blind testing programs in 2009. These rates were down from 2002, when more than 20% percent of forensic laboratories reported conducting blind tests [13,14]. Blind testing programs are concentrated in federal labs. In 2014, 39% of federal laboratories conducted blind tests, compared to between 5–8% of state, county, and municipal labs [13].

A landmark report from the National Academy of Sciences in 2009 recommended that forensic proficiency testing programs include blind tests where appropriate [15]. In 2016, the National Commission on Forensic Science recommendation to the Attorney General recommended that all Department of Justice Forensic Science Service Providers “seek proficiency testing programs that provide sufficiently rigorous samples that are representative of the challenges of forensic casework.” [16] Also in 2016, the President’s Council of Advisors for Science and Technology issued a stronger recommendation, noting that “PCAST believes that test-blind proficiency testing of forensic examiners should be vigorously pursued, with the expectation that it should be in wide use, at least in large laboratories, within the next five years. However, PCAST believes that it is not yet realistic to require test-blind proficiency testing because the procedures for test-blind proficiency tests have not yet been designed and evaluated.” [17]. This echoed concerns about feasibility made by the Forensic National DNA Review Panel more than a decade earlier [1]. We argue that the time has come to

¹ In other industries, proficiency testing may be required as part of individual certification or licensing programs. For an early review of the feasibility of blind testing for DNA laboratories as well as regulating legislation in other industries, see [1].

address these logistical challenges.

These reports highlight another benefit of blind testing: because blind tests must be similar enough to real cases enough to convince an examiner, they are effectively required to be similar to casework. This requirement is valuable not only for quality assurance perspective but also for outside consumers of proficiency test results, which can be cited in legal proceedings. A 2019 study showed that potential jurors shown a mock case gave a different likelihood the defendant left fingerprints at a crime scene when they were shown different proficiency test results for the fingerprint examiner in the case [18].

2. The meeting

In October 2018, the Center for Statistics and Applications in Forensic Evidence (CSAFE) hosted a meeting in Pittsburgh, at the Allegheny County Office of the Medical Examiner (ACOME), with laboratory directors and quality managers from county and state laboratories in the eastern US and the Houston Forensic Science Center. This meeting sought to understand the level of interest in implementing blinding, and particularly blind proficiency testing, and to identify the obstacles to implementing blind testing that laboratory managers could expect to encounter. Participants were invited based on the recommendation of CSAFE investigators and advisory board members, ACOME laboratory management, and the Association of Forensic Quality Assurance Managers (AFQAM), and included representatives from seven forensic laboratory systems ranging in size from a single laboratory with fewer than 50 employees to a seven-laboratory system with over 200 employees. Two of the quality managers were representing AFQAM. Additionally, professors, graduate students, and a post-doctoral researcher from three universities attended, representing fields ranging from statistics to psychology.

Peter Stout, CEO of the Houston Forensic Science Center,² provided a keynote talk, and he and two quality assurance managers from the Houston lab discussed their experience implementing blind proficiency testing at HFSC. Their blind proficiency testing program began in 2015 and now is the most robust blind proficiency program we are aware of in a non-federal forensic laboratory, with blind tests operational in the following divisions: biology, digital forensics, forensic multimedia, latent print comparison, latent print processing, toxicology, and seized drugs. In the firearms division, HFSC has implemented blind verification and quality control. In blind verification, two examiners (or more) analyze the same sample and produce a conclusion or report without knowing what the other examiner decided. This is similar to blind proficiency testing in that a previous conclusion is unknown to the verifying examiner. However, as ground truth is unknown, blind verification tests the consistency of the examiners' results; it does not provide information on whether the conclusion is accurate.

The laboratory has a goal of conducting a number of blind tests equivalent to 5% of casework in each discipline. Further, the laboratory publishes the results of all tests on a public-facing website. A healthy discussion followed the HFSC presentation, and other laboratories identified obstacles they had encountered and if and how they overcame these obstacles. There are significant logistical and cultural challenges in implementing blind proficiency tests in the forensic context, and many of the participating individuals came from labs that had not implemented blind testing at that time.

In the next section, we synthesize the results of the 1.5-day colloquium into a set of core challenges to implementation of blind

proficiency testing and the practices that have been used to successfully address those challenges. In addition, we identify the stakeholder community, that must be engaged for each challenge, as well as issues that are common across labs or that differ by laboratory size, type, or the discipline under consideration. Finally, we present a set of actions that could move the field forward.

A description of the November 2018 meeting on blinding can be found in [6]. Here, we synthesize the key findings related to challenges and potential solutions. The meeting included many substantive discussions. Many of the discussions led to consensus, but also areas of disagreement. The set of issues we describe below and recommendations was derived from these conversations and additional follow up discussions. We are deeply indebted to all who attended and shared their interests, concerns, and solutions. We have made every attempt to report consensus views and findings where possible. However, consensus was not reached on every issue; the views presented here are our own and should not be construed to represent all attendees.

The goal of our ongoing work is to better understand challenges to the implementation of blind proficiency testing and to develop solutions. (See Table 1) Straightforward solutions identified as a result of the meeting are provided in the next section. However, some challenges did not have solutions that are immediately implementable at labs of all types, and we provide discussion of steps needed to move forward with these hard problems.

3. Key challenges and solutions/next steps

3.1. Challenge – creating realistic test cases

A key challenge to implementing blind proficiency tests is that crime scene evidence can be complex to mimic. While for some disciplines such as toxicology, materials can be purchased from vendors such as RTI, for others, such as arson or ballistics or latent print analysis, evidence will need to be generated by the laboratory staff from locally purchased materials. Those who have successfully implemented blind proficiency testing also report that the detail-oriented nature of their examiners can present challenges. For example, at HFSC, a crowbar submitted, in a blind proficiency test for latent print processing was spotted because, the examiner told them, no one holds a crowbar in the way that would leave the print pattern found on the submitted item.

3.2. Solutions/next steps

Appropriate expertise must be developed such that the individuals preparing the test evidence understand the characteristics of evidence retrieved from a crime scene (e.g., how latent prints are collected and what constitutes a useable print). Ideally, an examiner from a discipline would aid in evidence preparation, but this could lead to potential un-blinding of test cases. If tests are prepared by quality assurance managers, they will learn over time as blind tests are discovered by examiners. Options discussed were the development of training programs, meetings (that could be attached to the annual meeting of the Association of Forensic Quality Assurance Managers or American Society of Crime Laboratory Directors), or a consortium of laboratories doing blind testing that could develop a shared evidence bank and lessons-learned compendium.

3.3. Challenge – creating realistic submission materials

As with test material preparation, submission materials (e.g., evidence documentation, laboratory request forms, crimes scene notes, etc) must look realistic. In addition to knowing a great deal

² We sometimes refer to the Houston Forensic Science Center as HFSC or as the Houston lab.

Table 1
Key challenges identified at meeting.

Test Creation	Test Management	Culture
Creating realistic test cases	Test submission to lab	100% accuracy myth
Creating realistic submission materials	Tracking test cases	Fear of failure/repercussions
Cost	Ensuring results are not released as real cases	
	Determining metrics to collect and report	

about the types of material the analyze, examiners know the patterns of cases and submissions they see. In early blind proficiency tests at HFSC, examiners spotted test cases because the handwriting on the submission form was “too neat.” Additionally, examiners know which kinds of drugs are likely to come from which neighborhoods or if certain officers only do large busts. Making these submission material seem realistic turns out to be as important as making the evidentiary sample seem realistic. This means that if submission materials commonly have messy handwriting and spelling mistakes, the test cases should have messy handwriting and spelling mistakes. Tests should come from jurisdictions that have reasonable rates of the type of crime that produces that type of evidence, and test cases should make sense given known patterns of individual law enforcement officers.

3.4. *Solutions/next steps*

This knowledge must be developed locally by staff managing the blind QA program in conjunction with the submitting agency.

3.5. *Challenge – cost*

Because declared proficiency tests are well established and perceived to provide valuable information, blind proficiency tests will probably need to be implemented in addition to current declared proficiency testing programs, not as a replacement. This reality presents a financial challenge. Stout estimated that HFSC spends as much on its blind proficiency testing program as it does on its traditional declared proficiency testing program. Most labs are not able to immediately double their proficiency testing budget.

3.6. *Solutions/next steps*

Laboratory managers indicated that while piloting blind testing might be feasible without significant budget increases, scaling a program requires a champion in the laboratory’s parent agency. Another potential way to address this financial challenge is for interested laboratories to form a consortium that facilitates sharing of materials and economies of scale with joint purchases. Also, if external proficiency test providers were able to develop blind proficiency tests (or materials for these tests), this could aid in making materials affordable.

3.7. *Challenge – test submission to lab*

In addition to cases looking realistic, test cases must be submitted by an outside law enforcement agency. For a laboratory that supports one jurisdiction, this means developing a relationship with that agency to support submission. For example, the HFSC supports the Houston Police Department, and HPD officials are involved in submitting the blind test cases. However, some laboratories support a number of jurisdictions. For example, the Allegheny County Office of the Medical Examiner, which houses the forensic laboratory for Allegheny County, processes evidence for more than 100 law enforcement agencies. Additional challenges occur in jurisdictions where examiners consult directly with law

enforcement. Further, if the laboratory staff has strong personal relationships with submitting agency personnel, loyalty and friendships may induce conversations a long the lines of, “You didn’t here this from me, but this laboratory request is special ...”

3.8. *Solutions/next steps*

The decision of which law enforcement agency or agencies to ask to serve as a submission partner should depend on the relationship between laboratory management and the LE agency (for example, likelihood of support for the program and likelihood that the information on which cases are tests can be contained and not leaked), and the patterns of crimes in the jurisdiction covered by the LE agency. Ideally, test submission would occur in a context where examiners are expected not to discuss work in progress with law enforcement. If such communication is an existing norm, it will need to be considered in case development and submissions. If a laboratory employs case managers to receive submitted cases and provide only necessary information to examiners, this can aid in ensuring examiners are not accidentally (or intentionally) alerted to blind tests.

3.9. *Challenge – tracking test cases*

Ideally, a blind proficiency test case would be flagged in a Laboratory Information Management System (LIMS) so the QA staff can keep track of the test and the examiner can use the LIMS as he or she does with everyday casework; however, not all LIMS are set up to easily offer a flag that is only visible to QA staff.

3.10. *Solutions/next steps*

Laboratories can consider choosing a LIMS on the basis of this feature availability or develop an in-house system to track test cases.

3.11. *Challenge – ensuring test results are not released as real cases*

As the goal of blind testing is to test the entire evidence processing pipeline, ideally the test should continue through to final report preparation. However, the reports should remain internal to the laboratory and not be released to the submitting agency or District Attorney’s office. Similarly, laboratories will need processes to ensure that test fingerprints or DNA profiles (which may, for example, come from QA staff) are not uploaded to national database systems like NDIS or NGI.

3.12. *Solutions/next steps*

To ensure reports are not released as if the test were a real case, the QA team will need to work with individuals in other units of the laboratory. The key issue is keeping the number of individuals who know that a case is a test as small as possible while ensuring results are not inadvertently released. HFSC deals with database concerns by generating dummy profiles in local systems that are not uploaded to state or national databases. In some jurisdictions, it may be

useful to have a contact in the district attorney's office as well as the submitting police agency. The exact number and makeup of the group that is "in the know" will depend on the size and structure of the laboratory and the context in which it operates.

3.13. Challenge – determining metrics to collect and report

Blind proficiency tests allow quality managers to determine if the right answer was obtained, and also if the correct process was followed. At the same time, laboratory staff must determine whether to include proficiency tests in overall metrics for the agency such as turn-around time and case submission number.

3.14. Solutions/next steps

For now, decisions about metrics will be made on a lab-by-lab basis. The development of a consortium could aid in standardization across labs. Additionally, engagement of the Association of Crime Laboratory Directors (ASCLD), Association of Forensic Quality Assurance Managers (AFQAM), or the Organization of Scientific Area Committees (OSAC) could help identify standard measures. HFSC staff noted that they publish test results on their public-facing website, but many attendees felt this level of openness might not be possible in their jurisdictions. It is not necessary to make test results public to implement a useful program. Where and how to report results is a decision that can be made on a lab-by-lab basis.

3.15. Challenge – myth of 100% accuracy

The idea that errors occur in forensic analyses pushes against a cultural history in the field. In many disciplines, examiners have historically claimed the ability to generate perfect identifications or error rates of zero [19,20]. No scientific discipline actually achieves error rates of zero, and errors have been found in many disciplines in forensics [21], and even in declared proficiency tests [22]. Blind proficiency tests can flag errors or mistakes related directly to examiner performance and also with the laboratory pipeline, and are one of the only methods that can identify misconduct. Further, they can help identify issues related to the technique itself. (See the discussion of conforming and non-conforming work above on how even conforming work can produce inaccurate results.) Each of these findings presents a different quality management issue, and they are all important to understand and address. However, implementing changes based on the argument that the new process will find more mistakes pushes against a cultural history in the field, and is complicated by the fact that laboratories may be compelled to provide their test results in legal proceedings. The perceived potential for long term consequences of a problematic blind proficiency test is a barrier to participation for many lab directors.

3.16. Solutions/next steps

All present at proficiency testing meeting the meeting emphasized that it would only be possible to implement blind proficiency testing with active support from senior management—essentially, a champion somewhere near the top of the reporting structure. Stout noted that laboratory directors need to communicate with their supervisors and stakeholder about the cost, both financial and reputational, of nonconforming work and inaccurate results in real cases. Further, laboratory directors must embrace a culture of continuous improvement where discovering problems is a good, not bad, thing. Other industries have adopted this mindset. Forensic laboratories could benefit from looking to aviation and the way airline simulators are used; the airline community does not treat

simulated crashes as if they were real crashes, but rather as opportunities to learn and improve.

4. Moving the field forward

At the meeting, the practitioner participants indicated that interest in blind proficiency testing was more widespread than the academic researchers had previously realized. However, at the present time, laboratories are each making decisions about whether to implement blind testing, and navigating the challenges that arise, independently. All participants emphasized that the financial costs, logistical requirements, and perceived risk of implementing a blind proficiency testing program are too great for a QA manager or discipline chief to implement their own; the effort requires a senior champion. Finding these senior champions is complicated by the fact that a small minority of state and local laboratories are led by a scientist. In the majority of state and local crime laboratories, labs are headed by or report up to law enforcement officials. The discussion indicated that when laboratories report up to a person with a scientific background, it is usually less challenging to argue the value of blind testing than in situations where laboratories report into a police department. In fact, HFSC, which is a champion of blind testing, and ACOME, which hosted the meeting, are examples of the laboratories that are headed by scientists. (HFSC CEO Peter Stout is a PhD chemist; in Allegheny County, the forensic laboratory is housed in the county medical examiner's office, which is headed by a board-certified forensic pathologist.)

However, a number of other stakeholders could provide momentum for the field as a whole (see Table 2 for a list of stakeholders). The participants suggested proposing workshops at meetings of the Association of Crime Lab Directors (ASCLD) and the Association of Forensic Quality Assurance Managers (AFQAM), which could serve to broaden the discussion. Additionally, increasing engagement the Organization of Scientific Area Committees (OSAC) standards development process could provide momentum. Obviously, if accreditation bodies could be persuaded to make blind proficiency testing accreditation requirement, laboratory directors would have a mandate to pursue the resources to retain maintain their accreditation. Ideally, blind proficiency testing will become a marker of a strong QA program and high-quality forensic science. This culture change would lead to increased adoption until blind proficiency testing is the norm.

Commercial proficiency test providers were seen as another key audience to engage. While all laboratories conducting blind testing were anticipating producing their own blind test samples in the near term, because of the logistical concerns outlined above, it could be possible for commercial vendors to provide materials that could be integrated into in-house tests or to eventually develop full blind tests for some disciplines. Additionally, laboratory directors noted the need for appropriate features in the Laboratory Information Management System (LIMS) to support blind testing, including a flag only visible to select staff that could label a case a blind test.

Laboratory representatives at the meeting felt that even in initial stages, increased communication between interested laboratories and ideally a formal collaboration infrastructure could be helpful. Formalizing the network of laboratories interested in blind proficiency testing through a consortium and developing regular communication channels could provide a mechanism for sharing best practices, strategies and possibly a test bank of materials. Additionally, as more laboratories move forward with this model, it will be important to assess the results. Each laboratory will be producing test results for its own purposes, but a centralized data repository connected to academic researchers could help

Table 2
Key stakeholders.

Laboratories	Clients	Professional Bodies	External Organizations
Examiners	Law enforcement	Association of Forensic Quality Assurance Managers	Accreditation bodies
Quality assurance staff	Prosecutors	Association of Crime Laboratory Directors	Proficiency test providers
Laboratory management	Defense attorneys	Organization of Scientific Area Committees	Researchers at CSAFE, RTI, and other academic institutions
	Judges	American Academy of Forensic Sciences	
	Jurors		

determine standard metrics and conduct meta-analyses and inter-laboratory studies. Such a consortium would also provide support to QA managers who are in discussions with their senior leaders by producing communication tools that highlight the importance of blind proficiency testing being implemented by high quality laboratories.

5. Discussion

Blind proficiency testing is an important tool to add to forensic quality assurance programs. Blind tests can demonstrate quality and provide insight into the errors that occur in a laboratory in everyday casework. Unlike declared proficiency tests, with blind testing examiners believe they are analyzing real casework, and the test samples resemble everyday casework. Therefore, when implemented properly, this type of testing avoids the bias that arises from examiners knowing they are being tested, and from having tests that are too easy or are simply different from what the examiner sees in real cases. However, there are challenges to the implementation of blind proficiency testing in forensic laboratories, especially small laboratories because of the requirements in personnel and costs. A meeting held in Pittsburgh in 2018 led to the distillation of some of these challenges, as well as some solutions and next steps.

The meeting highlighted challenges to increasing implementation of blind testing are significant, considering the ways in which forensic laboratories differ from testing laboratories in other disciplines. However, there was general agreement that the potential benefits of implementing blind proficiency testing more widely would justify the investment of resources required to make it a standard component of forensic quality assurance programs, with continued analysis to assess results and impact as implementation increases.

Blind proficiency tests are not error rates studies, nor do we suggest they replace focused error rates studies. However, in aggregate, the data from widespread implementation of blind proficiency tests could improve understanding of both individual performance and errors in forensic techniques as applied in practice, and if and how performance varies between disciplines. Implementing blind proficiency testing has the potential to enable better accountability for forensic laboratories, and it could lead to reducing errors in forensic science.

Funding declaration

This work was funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln. Note: Robin Mejia and Maria Cuellar both receive funding from CSAFE through their respective institution's subawards with Iowa State. Jeff Salyards is a

member of the CSAFE strategic advisory board, a volunteer position.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

References

- [1] J.L. Peterson, G. Lin, M. Ho, Y. Chen, R.E. Gaensslen, The feasibility of external blind DNA proficiency testing. I. Background and findings, *J. Forensic Sci.* 48 (2003) 2002042, <https://doi.org/10.1520/JFS2002042>.
- [2] L.C. LaMotte, G.O. Guerrant, S.D. Lewis, C.T. Hall, Comparison of laboratory performance with blind and mail-distributed proficiency testing samples, *Publ. Health Rep.* 7 (1977).
- [3] T.M. Stull, Variation in proficiency testing performance by testing site, *J. Am. Med. Assoc.* 279 (1998) 463, <https://doi.org/10.1001/jama.279.6.463>.
- [4] G. Whitman, R. Koppl, Rational bias in forensic science, *Law, Probability & Risk* 9 (2010) 69–90, <https://doi.org/10.1093/lpr/mgp028>.
- [5] B. Budowle, M.C. Bottrell, S.G. Bunch, R. Fram, D. Harrison, S. Meagher, et al., A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement, *J. Forensic Sci.* 54 (2009) 798–809, <https://doi.org/10.1111/j.1556-4029.2009.01081.x>.
- [6] R. Mejia, M. Cuellar, J. Salyards, *Blinding at Forensic Laboratories: A Meeting Report*, 2019.
- [7] B.O. Gardner, S. Kelley, K.D.H. Pan, Latent print proficiency testing: an examination of test respondents, test-taking procedures, and test characteristics, *J. Forensic Sci.* (2019), <https://doi.org/10.1111/1556-4029.14187>.
- [8] A.J. Koertner, H.J. Swofford, Comparison of latent print proficiency tests with latent prints obtained in routine casework using automated and objective quality metrics, *J. Forensic Ident.* 68 (2018) 379–388.
- [9] B. Max, J. Cavise, R.E. Gutierrez, Assessing latent print proficiency tests: lofty aims, straightforward samples, and the implications of nonexpert performance, *J. Forensic Ident.* 69 (2019) 281–298.
- [10] G.S. Cembrowski, R.E. Vanderlinde, Survey of special practices associated with College of American Pathologists proficiency testing in the Commonwealth of Pennsylvania, *Arch. Pathol. Lab Med.* 112 (1988) 374–376.
- [11] D.C. Murrie, B.O. Gardner, S. Kelley, I.E. Dror, Perceptions and estimates of error rates in forensic science: a survey of forensic analysts, *Forensic Sci. Int.* 302 (2019) 109887, <https://doi.org/10.1016/j.forsciint.2019.109887>.
- [12] P.J. Parsons, A.A. Reilly, D. Esernio-Jenssen, L.N. Werk, H.C. Mofenson, N.V. Stanton, et al., Evaluation of blood lead proficiency testing: comparison of open and blind paradigms, *Clin. Chem.* 9 (2001).
- [13] A.M. Burch, M.R. Durose, *Publicly Funded Forensic Crime Laboratories: Quality Assurance Practices*, Bureau of Justice Statistics 2016, 2014, p. 11.
- [14] M.R. Durose, B. Statistician, K.A. Walsh, A.M. Burch, *Census of Publicly Funded Forensic Crime Laboratories*, Bureau of Justice Statistics 2012, 2009, p. 14.
- [15] National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward*, National Academies Press, Washington, D.C., 2009, <https://doi.org/10.17226/12589>.
- [16] National Commission on Forensic Science, *Recommendation to the Attorney General Proficiency Testing*, 2, 2016.
- [17] Presidents Council of Advisers on Science and Technology, *REPORT to the PRESIDENT Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 2016.
- [18] G. Mitchell, B.L. Garrett, The impact of proficiency testing information and error aversions on the weight given to fingerprint evidence, *Behav. Sci. Law* 37 (2019) 195–210, <https://doi.org/10.1002/bsl.2402>.
- [19] S.A. Cole, *More than Zero: Accounting for Error in Latent Fingerprint Identification*, 95, 2005, p. 95.
- [20] S.A. Cole, *Suspect Identities: A History of Fingerprinting and Criminal Identification*, Harvard University Press, 2009.
- [21] M.J. Saks, J.J. Koehler, The coming paradigm shift in forensic identification science, *Science* 309 (2005) 892–895, <https://doi.org/10.1126/science.1111565>.
- [22] L. Wilson-Wilde, S. Smith, E. Bruenisholz, The analysis of Australian proficiency test data over a ten-year period, *Forensic Sci. Pol. Manag.* 8 (2017) 55–63, <https://doi.org/10.1080/19409044.2017.1352054>.