Deadline for Comments: 8/26/2019
ASB Standard 063: Implementation of 3D Technologies in Forensic Firearm and Toolmark Comparison Laboratories, First Edition

Note: a specific Proposed Resolution must accompany each comment or it cannot be considered.

#	Section	Type of Comment (E-Editorial, T-Technical)	Comments	Proposed Resolution	Final Resolution
1	4.1.1	T	What is an "organization with appropriate knowledge and/or expertise" for developmental validation? The absence of specifics renders this standard a non-standard and yet this is a critical aspect of sound developmental validation	Define appropriate knowledge and expertise and submit the definition for public comment.	Reject: Based on ISO 17025 the onus is on the management to determine suitable qualifications for laboratory work. Therefore, the management determines appropriate knowledge and expertise.
2	4.1.3.4	T	Peer-review and the subsequent larger post-publication review by the scientific community is the hallmark of sound science. In the absence of a regulatory body, normal market forces, and the multi-disciplinary nature (e.g. statistics, computer science, research design) of the field of forensic science, rigorous and transparent developmental validation is an essential component of the development of accurate and reliable comparison software. Dissemination of validation data by means of alternatives such as presentation at a scientific meeting—which is often presented at the 10,000 foot level and without the full data—should not supplant peer review in a widely available scientific journal.	Include a requirement for peer reviewed publication in a widely available journal, and delete the parenthetical about other means of dissemination.	Accept with modification: The text was modified to clarify that even a presentation would have gone through a peer-reviewed process.
3	4.1.3.6	T	If the process outlined in the section is used then the positions of those involved and third party reviewer should be documented and that documentation should be made available to end users (e.g. courts, prosecutors, defendants).	Add language requiring that the positions of those involved and third party reviewer should be documented and that documentation should be made available to end users (e.g. courts, prosecutors, defendants).	Reject with modification: Sections 4.1.2.3 and 4.1.3.2 were modified to provide further clarity regarding documentation. Documentation availability is a laboratory specific decision that is outside the scope of this document.
4	4.2	T	Language should be added requiring that data and records generated during deployment validation should be maintained and organized so that a qualified independent expert can evaluate the testing and results from the deployment validation and assess the technology's or procedure's ongoing performance.	A section should be added to 4.2 requiring that data both from the deployment validation and from all subsequent validation updates and performance checks be maintained and be readily available upon request.	Reject with modification: Sections 4.2.2.1 and 4.2.3.2 were modified to provide further clarity regarding documentation. All levels of validation shall be documented activities. Availability of these documents is a laboratory specific decision that is outside the scope of this document.
5	4.3	T	Language should be added requiring that data and records generated during deployment validation should be maintained and organized so that a qualified independent expert can evaluate the testing and results from the deployment validation and assess the technology's or procedure's ongoing performance.	A section should be added to 4.3 requiring that data both from the deployment validation and from all subsequent validation updates and performance checks be maintained and be readily available upon request.	Reject: Documentation availability is a laboratory specific decision that is outside the scope of this document.
6	4.5	T	This section should include a requirement that the testing performed include samples that mirror the most difficult case work likely to be encountered. Just as one simulates dangerous or difficult conditions when testing other high risk professions (e.g. pilots, EMT's) examiners should also be expected to perform well on tests that reflect the most difficult challenges they are likely to encounter in case work. Likewise, as is done in other industries (e.g. medicine), the test taker should not know that they are being tested to ensure that the test subject is using the same procedure and decision making thresholds as they would use in actual case work.	Tests should be comprised of problems that reflect the most challenging problems encountered in case work, including damaged and other low quality samples. To the extent possible the test subjects should not know they are being tested	Reject with modification: It is not feasible for competency testing to be blind. However an edit was made to section 4.6. to address this for proficiency testing.
7			Bibliography citation format may need some revision, esp. 1].		Accept
8			My comments, as they have previously, apply to all three topography and comparison software related standards (61-63). While all three standards clearly represent an admirable first step towards greater objectivity in the field of firearms examination, one clearly earned through concerted and well-intentioned labor, a wealth of lingering concerns nevertheless prevents me from voting in favor of any of the three. My main points of contention (though not all encompassing of my objections) are as follows:		This comment was broken down into 6 sections as following and a resolution for each section has been provided.

#	Section	Type of Comment (E-Editorial, T-Technical)	Comments	Proposed Resolution	Final Resolution
8 (8.1)			(1) Software Engineering Standards: Outside of the forensic sphere, the Institute of Electrical and Electronics Engineers (IEEE) has developed a whole series of Standards meant to define best practices for the development of software systems, most relevant among them being IEEE, "Std. 2012-2016 Standard for System, Software, and Hardware Verification and Validation," (2016). And this body would be unwise to ignore the consensus guidelines of the world's largest and most prestigious software body. In fact, multiple groups who have addressed probabilistic genotyping systems in the DNA sphere have arrived at precisely that conclusion. See United Kingdom Forensic Science Regulator, "Guidance: DNA Mixture Interpretation Software: Validation, FSR-G-223," (2017); M.D. Coble et al., "DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications," 25 For. Sci. Int'l Genetics 191, 192 (2016); Nathaniel Adams et al., "Letter to the Editor- Appropriate Standards for Verification and Validation of Probabilistic Genotyping Systems," 63 J. For. Sci. 339 (2018). Yet, despite the significant problems caused by the DNA community's late adoption of best practices for software, Standards 61-63 make no effort to track IEEE's guidance.		Reject (1): The IEEE standards are intended for developers while these standards are intended for end users. The software described in these documents can be empirically tested on real-world data. Software performance is important and the document describes three stages of validation testing to ensure that the software meets the needs of the end user.
8 (8.2)			(2) Independence: One of the gaps between IEEE and Standards 61-63 actually brings me to my second concern, that of independent validation. Developers are by definition self-interested and until these Standards require full developmental style validation occur by independent researchers they fall short of what the legal and forensic communities deserve.		Reject (2): This falls outside of the scope of this document as these requirements are for the end user and not the developer. The end user must validate the technology prior to use. The strength of the validation is a function of the scale/scope and quality of the validation completed. Larger more independent validations are more substantial than smaller or less independent validations.
8 (8.3)			(3) Laboratory Validation: To the extent laboratory implementation validation might plug this gap, however, Standards 61-63 do not lay 0out rigorous enough requirements, most centrally because they offer little in the way of the qualifications required of those performing such work. Are we really to believe that a science bachelors degree suffices for lab personnel necessarily to have the background in statistics, computing, and experimental design necessary to catch flaws in these new systems?		Reject (3): The strength of a validation is affected by several factors, these include scale/scope as well as who conducts the validation. The documents state that development validation shall be peer reviewed and publicly available which means that one or more peers would provide an independent check. The document states that development and deployment validations shall also be evaluated by a technical reviewer. The document states that validation must be performed by someone with appropriate knowledge or expertise. Laboratory personnel are self-motivated to find someone appropriate. Otherwise the laboratory may have problems down the road.
8 (8.4)			(4) Transparency: Whatever a company might be able to claim in the private sphere in terms of patents, trade secrets, etc… such proprietary incentives are a poor fit for the criminal justice system. While it might* go to far to demand open source programs, these standards do not go as far even as has the DNA community. It should not be enough to publish the scientific principles of a system. At minimum all algorithms used must be public and peer-reviewed as well. Every major probabilistic genotyping system has managed at least that much, and the firearms examination community cannot afford to do less.		Reject (4): This falls outside of the scope of this document as these requirements are for the end user and not the developer. The document states peer-reviewed publication of the scientific principles shall be required. As described above, firearm and toolmark examination algorithms can be tested empirically using large real-world test sets that are representative of actual casework. The strength of the validation is a function of the scale/scope and quality of the validation completed.
8 (8.5)			(5) Reporting Language: Standard 62's treatment of reporting results is disturbing and likely to cause real difficulty in the courts. In allows for use of the word match when describing Category 0 scores in the false assurance that the fairly confusing statement about statistical significance that follows will cure the prejudice associated with the word match. And it allows analysts to testify to "high" match scores in Category 1 as if such statements will not be interpreted incorrectly by jurors. Even under category 3 it does not require any explanation to jurors of the underlying probability and statistical principles associated with the program (is it Bayseian, frequentist, do they need to take into account a prior probability or base rates???)		Reject (5): Does not apply to document 063.

#	Section	Type of Comment (E-Editorial, T-Technical)	Comments	Proposed Resolution	Final Resolution
8 (8.6)			(6) Validation Samples: Especially considering the massive role left to labs to validate and implement the systems described in Standards 61-63, those documents provide grossly insufficient guidance as to the types of samples necessary to particular validation studies opening these technologies up to abuse. While it may not be possible to exhaustively list all possible sample variations, these standards can at least set a floor or minimum and describe what they are doing as such. That would allow researchers to do more but would not allow labs or developers to get away with doing less (ie failing to run any known subclass samples).		Reject (6): Does not apply to document 063.
9			As with my previous comments, I am writing with respect to the three standards, and as before I believe these standards are a very good, thoughtful start but need more fleshing out to give guidance to practitioners. 1) I agree with Richard about the failure to cite to or address the IEEE standards. Frankly I don't know enough about contents of the IEEE validation standard to know what the precise discrepancies are, but I believe it should be treated as a normative reference (sec. 2) for all three standards. 2) Who should conduct validation: a. For developmental validation, the documents say it should be conducted "by an organization with appropriate knowledge and/or expertise." What does this mean? What constitutes "appropriate" knowledge and/or expertise? This should be more precisely defined. Also, studies by organizations/instituti-ons independent of the developer should be required. b. For deployment validation, the docs say it should be conducted by someone with minimum of a bachelor's degree w/ a science major. If the requirements of deployment validation were set out in great detail, I might not be concerned with this fairly low bar. However, this is not the case. Instead, four one-line "aspects [that] shall be documented" are set forth, and otherwise the persons/entities conducting the study are given unfettered discretion on study design. As I've said before, I think the validation aspect of these documents needs to be bulked up significantly to give practitioners meaningful guidance. For example, there is no mention at all of concepts of sensitivity or specificity, or of the idea of designated inconclusive ranges—all concepts essential to the actual deployment of this technology/software. If additional clarifying detail regarding the design of validation studies is not going to be added to this document, I would at very least make it clear these documents do not provide a formula for adequate validation, and require the validator to consult with a statistician or someone with expertise in study design. 3) Sample sets for validation: As I mentioned in my last set of comments, I don't think these documents provide sufficiently specific guidance about what kinds of samples have to be included at minimum to make the studies adequate. Certainly, the example given in Annex A should include samples that test the limits of the system, e.g. damaged and poorly marked ammunition, toolmarks left by consecutively manufactured firearms, etc. 4) Preservation and disclosure of data: I'd like to see some language in these documents emphasizing the importance of independent review of data, and requiring validators (whether involved with developmental, deployment, or performance checks) to maintain data and make it		Reject (1): Please see above resolution of comment #8 part 1. Reject (2a): Appropriate expertise is application specific. Reject (2b): The quality of the validation is related to the quality of those conducting the validation and the study design; the comment that a formula is not presented is already covered, the need to seek out those with appropriate expertise is explicitly mentioned. Reject (3): Does not apply to this document. Reject (4) Standard mentions documentation and peer review; documentation preservation should follow lab protocols for all other laboratory equipment; no need to restate here. Reject (5): Does not apply to this document.
10			In light of the objections raised by Richard Gutierrez, I would urge further study regarding the discrepancies between these Standards and the Standards of the IEEE.		Reject: No text changed proposed. The IEEE standards are intended for developers while these standards are intended for end users.
11			Under 4.1.3 there should be an alternative for a reviewer for sole practitioners.		Reject: Development and deployment validation need that technical review.