Response from the Physical Fit Task Group to the Scientific & Technical Review Panel Final Report for 2022-S-0015

Standard Guide for Forensic Physical Fit Examination

Organization of Scientific Area Committees (OSAC) for Forensic Science

August 2022

Response to STRP Final Report 2022-S-0015

Organization of Scientific Area Committees (OSAC) for Forensics Science, Trace/Materials Subcommittee, Physical Fits Task Group
August 2022

.

The Trace/Materials' Physical Fits Task Group evaluated the recommendations of the Final STRP report 2022-S-0015, and the comments are provided below, highlighted in color blue. We thank all the STRP members for the time and recommendations provided.

Physical Fit Task Group Members:
Kris Gates
Susan Gross
Tammy Jergovich
Jenny Lounsbury
Daniel Mabel
Andria Mehltretter
Michelle Mercer
David Northrop
Troy Nowak
Meghan Prusinowski
Candie Shegogue
Richard Thomas
Tatiana Trejos

Report Summary:

The STRP has reviewed and discussed this draft standard and has formed consensus that many sections below have been addressed, however, there remain notable weaknesses in the Human Factors, Quality Assurance, and Reporting Results sections. Ideally, these weaknesses should be addressed before this document moves forward as they will continue to be problematic as this document moves through the SDO process. If this is not possible, they should be addressed in future iterations of this document as additional data and supporting information becomes available.

The Scientific and Technical Review Panel (STRP) for "Standard Guide for Forensic Physical Fit Examination" is an independent panel appointed by the National Institute of Standards and Technology (NIST). A STRP is established with a range of experts to consider how well a standard meets the needs of the forensic science, law enforcement, and legal communities, and to recommend improvements to the standards under review. The STRP appreciates the efforts of Tatiana Trejos, Trace Materials Subcommittee affiliate, while serving as the subcommittee liaison to this STRP during the review process.

The STRP began its review process with a kickoff meeting on December 8, 2021 and concluded with this STRP final report. The panel reviewed the draft standard and prepared comments for the Trace Materials Subcommittee.

R/ The Physical Fit Task group met on July 20, 2022 to evaluate the comments provided in this report, and we arrived at a consensus on how to address those recommendations in a revised document that is provided with the tracked changes.

Report Components:

The STRP reviewed this draft standard against OSAC's *STRP Instructions for Review* which include the following content areas: scientific and technical merit, human factors, quality assurance, scope and purpose, terminology, method description and reporting results. The details below contain a brief description of each reviewed content area and the STRP's assessment of how that content was addressed in the Draft OSAC Proposed Standard.

1. **Scientific and Technical Merit:** OSAC-approved standards must have strong scientific foundations so that the methods practitioners employ are scientifically valid, and the resulting claims are trustworthy. In addition, standards for methods or interpretation of results must include the expression and communication of the uncertainties in measurements or other results.

    1.1 Consensus View – Upon deliberation, the STRP reached a consensus that the document, in general, is technically sound. With respect to scientific and technical merit, the material presented is within the scope; the document is well cited with respect to the supporting scientific and technical literature; and the limitations of the guidance provided and the techniques are appropriately expressed.

2. Minority View – The references provided in support of the scientific and technical merit of the techniques do not rise to the level of validity desired and there is a lack of empirical work to support the validity of the technique and the illustrative report examples presented in sections 14.1.1 and 14.1.2. To this end, the conclusions drawn from the workflows presented herein could be potentially misleading.

R/ We appreciate the consensus view. Regarding the minority view, we believe the references provide support for the scientific validity of the guide. We have added the most recent empirical work that evaluates the scientific validity of physical fits, and reference #2 is a review manuscript that discusses in more detail the existing literature and the current state of the field. We do recognize that there are still some research gaps to be addressed, but a standard like this one is the right step toward standardized protocols that can assist the community in enhancing their practice.

3. **Human Factors:** All forensic science methods rely on human performance in acquiring, examining, reporting, and testifying to the results. In the examination phase, some standards rely heavily on human judgment, whereas others rely more on properly maintained and calibrated instruments and statistical analysis of data.

    2.1. Consensus View – There are two human factors issues that remain in the

document after revisions were made in response to comments from the public and the STRP: 1) the verification procedures are insufficient to safeguard against error as a result of bias and/or human judgment; and 2) the section addressing precautions to minimize bias does not provide enough guidance to help examiners and laboratories implement these mitigation strategies.

Verification Procedures

i. The first "Discussion" piece under 3.2.8 should be edited to reflect that the second examiner analyzes the same materials independently and their analyses can corroborate the initial examiner's judgment if they agree, or lead to a correction if they do *not* agree. The STRP suggests: "The process through which the analyses of the initial forensic examiner for a particular set of materials or evidence are compared to analyses completed by a second, independent examiner so that the findings of the first examiner are corroborated when there is agreement or can be corrected in situations where there is disagreement."

R/ We added wording to section 3.2.8.1 to address this recommendation. Also, we added clarification in section 15.

ii. The second "Discussion" piece under 3.2.8 states that verifications can be "open or blind". If the verifications are open, then they are not independent, which contradicts the first discussion point. Open verification permits the second examiner access to information about the judgments and thought processes of the first examiner, which prevents a truly independent evaluation of the evidence.

R/ We added a sentence to section 3.2.8. to address this concern. It reads now, "Blind verifications are more robust than open verifications". We used this additional language to provide further clarification and also align with other disciplines' standard language regarding blind verifications.

The committee wants to clarify to the STRP that when we use the term independent, it means a second examiner who did not partake in the original analysis and arrives at a conclusion that confirms or contradicts the first examiner's conclusion. This independence is maintained whether the verification is blind or open. In an open verification, the independent examiner has access to the overall conclusion (or the second examiner can infer that he/she is verifying a fit). In a blind verification, the second examiner does not have access to the first examiner's notes, results or conclusions. Fully blind verifications may not be possible in some jurisdictions or laboratory quality management systems that only verify "physical fits".

iii. Similar to the previous point, in 15.2 of the Standard Guide for Physical Fit Examination, the document states that a "review and examination of the actual evidentiary material or by reviewing the documentation of the physical fit" are both acceptable methods of verification, but the latter form of verification is not independent. In addition, the language in 15.2.1 suggests that the default is for verifiers to work with the documentation rather than the actual material. The default should be another, independent analyses of the actual evidence without access to other information unless that is not possible for some reason (e.g., the evidence had to be altered or manipulated to perform the first analysis).

R/ We have added an example of the documentation to clarify. The documentation that can be verified is, for example, an image of high quality of the compared items, so that the second examiner can evaluate if a fit is demonstrable, or not. However, the independence of conclusions is maintained in both verifications (on the actual evidentiary material or the images) because the second examiner is applying their own knowledge, skills and expertise.

1. *Suggested Alternative Wording*: Remove 15.2.1 and replace 15.2 with:*Verification should be in the form of review and examination of the actual evidentiary material. If the actual evidence cannot be observed by the verifier (e.g., the evidence had to be altered or manipulated to perform the analysis), then documentation which clearly and objectively depicts the physical fit may be used.*

R/ Section 15.2 was edited to clarify what documentation we mean.

2. *Potential Alternative Solution*: If open verification is something that the SAC and Physical Fit analyses community wishes to continue using, then an information management procedure should be used. First, the second analyst (who has not yet been involved in the case) should evaluate the original evidence or documentation that is as close to the original evidence as possible and make an independent determination about physical fit based on those materials only. Once that decision and any associated reasoning is recorded, the second examiner can get access to the other documentation produced by the first examiner when they made their decision about physical fit. If this information changes the judgment made by the first examiner, this should be documented, and the change should be explained and justified so that there is transparency regarding what information influenced each determination by the verifier.

R/ Section 15.2 was edited to clarify what documentation we mean.

Mitigating Bias Effects

i. In 9.1.4.1 and 9.1.4.2, it would be helpful to provide some citations that could serve as resources to help examiners learn about procedures and tools to help mitigate bias. Otherwise, this section does not provide enough guidance about what should be done to help reduce the effect of biasing information. Some examples of citations are provided below, however, the SAC can also refer to the internal Human Factors Task Group guidance document available to OSAC members:
   a. Quigley-McBride, A., Dror, I.E., Roy, T., Garrett, B.L., & Kukucka, J. (2022). A practical tool for information management in forensic decisions: User Linear Sequential Unmasking-Expanded (LSU-E) in casework. *Forensic Science International: Synergy, 4,* e100216. https://doi.org/10.1016/j.fsisyn.2022.100216 [Easy to access materials available here too: https://osf.io/xm3ru]
   b. Spellman, B.A., Eldridge, H., & Bieber, P. (2022). Challenges to

reasoning in forensic science decisions. *Forensic Science international: Synergy, 4*, e100200. https://doi.org/10.1016/j.fsisyn.2021.100200

c. Dror, I. E. (2020). Cognitive and human factors in expert decision making: Six fallacies and the eight sources of bias. *Analytical Chemistry*, *92*(12), 7998-8004. https://dx.doi.org/10.1021/acs.analchem.0c00704

R/ Added citations in 9.1.4 and the respective references in section 18

In 9.1.4.3, the wording is a bit confusing. The STRP suggests: "Assessing questioned samples separately prior to comparing those samples to the known samples." Note that assessing the questioned samples in this way first is not the same as "sample handling". It should also involve an assessment of the value and quality of the evidence, an evaluation of its features and whether they are particularly distinctive or useful in a physical fit analysis context, etc.

R/ The section was edited as suggested.

ii. In 9.1.4.5, the SAC should specify when/how often they expect technical review and verification to occur to mitigate bias effects. In addition, they are presenting verification as a solution to bias, but are permitting open verification. Because bias is difficult to spot and often occurs outside awareness, blind verification is the best way to protect against bias using verification procedures.

R/ The committee disagrees with this comment. The frequency and procedures for a technical review or verification are to be determined by each laboratory management and quality system and are typically covered in their quality manuals. The statements in this section are in compliance with laboratory accreditation norms as per ISO/IEC 17025 and ANAB 3125.

2.2. Minority View – The Human Factors elements of the techniques discussed in the guide are adequately addressed in the document.

4. **Quality Assurance:** Quality assurance covers a broad range of topics. For example, a method must include quality assurance procedures to ensure that sufficiently similar results will be obtained when the methodology is properly followed by different users in different facilities.

3.1. Consensus View – Quality assurance procedures should be well defined to minimize risk and to ensure consistency amongst laboratories. As written, the requirement for verification (15.1 – 15.3) is vague, which prevents consistent application of the requirement as well as introduces risk by allowing verifications to be performed with sketches and/or case documentation. Instead, it should require an independent examiner to evaluate the evidence, either through physical comparisons or with photographs that clearly and objectively depict the physical fit. In addition, review of the case documentation should be considered part of the technical review process and is not adequate for a verification.

R/ We have updated 15.2 to clarify that the documentation allowed for verifications are quality images, and not notes or sketches as interpreted by the STRP. The images need to clearly and objectively demonstrate the features so that a second examiner can independently reach a conclusion in their

physical fit examination.

3.2. Minority View – The Quality Assurance aspects provided in the document are sufficient for the purpose of this guide.

5. **Scope and Purpose:** Standards should have a short statement of their scope and purpose. They should list the topics that they address and the related topics that they do not address. Requirements, recommendations, or statements of what is permitted or prohibited do not belong in this section.

    4.1. Consensus View – The STRP finds that the information presented within the document is within the well-defined scope and conforms to the overall purpose of the guide.

    4.2. Minority View – None.

6. **Terminology:** Standards should define terms that have specialized meanings. Only rarely should they give a highly restricted or specialized meaning to a term in common use among the general public.

    5.1. Consensus View – The STRP finds that the draft standard defines appropriate terms with specialized meaning consistent with ASTM E1732 Standard Terminology Relating to Forensic Science and C1256: Standard Practice for Interpreting Glass Fracture Surface Features. The draft standard's reference to the OSAC Lexicon and ISO 17025:2017 provides consistency to relevant terms. The document balances the need for definitions while avoiding defining commonly used terminology. Examples are appropriately added to the various terminology such as in Section 3.2.

    5.2. Minority View – None.

7. **Method Description:** There is no rule as to the necessary level of detail in the description of the method. Some parts of the method may be performed in alternative ways without affecting the quality and consistency of the results. Standards should focus on standardizing steps that must be performed consistently across organizations to ensure equivalent results. Alternatively, standards can define specific performance criteria that are required to be demonstrated and met rather than specifying the exact way a task must be done. For example, it may be enough to specify the lower limit for detecting a substance without specifying the equipment or method for achieving this limit of detection.

    6.1. Consensus View – The STRP considers the method description provided in the guide to be fit for purpose and, despite some notable weaknesses in the area of Human Factors, would generally lead to sound conclusions if followed.

    6.2. Minority View – None.

8. **Reporting Results:** Methods must not only be well described, scientifically sound, and comprehensive but also lead to reported results that are within the scope of the standard, appropriately caveated, and not overreaching.

7.1. Consensus View – Upon deliberation of the STRP, it is the consensus of the group that the Reporting Results section needs clarification to resolve some ambiguities. Most specifically, exclusions based on the examinations performed through the guidance of this document should be explicitly stated.

The response to the feedback stated that there is a binary decision of "Physical Fit" or "No Physical Fit" in order to leave the door open for further analysis. Although this is a required option when class characteristics concur and no physical fit could be made, there will be situations in which an exclusion occurs based on observations made during the physical fit examination. In such instances, the door of ambiguity should be closed here and the exclusion should be reported at this point.

Additionally, the Report Wording Examples section needs better examples supported by empirical data. The examples provided here are somewhat ambiguous and could be problematic based on range of interpretations. To this end, and based on the comment on exclusions outlined above, there should be at least one example of an exclusionary conclusion presented in the document.

R/ We added an exclusion example on section 14.2.4
The references included in this document provide some empirical data and there is continued research in this area. It is not feasible to provide empirical data for all of the possibilities.

7.2. Minority View – The Reporting Results section is sufficient and fit for purpose.