



COMMENTS BY THE OSAC LEGAL RESOURCE COMMITTEE (LRC)

TO: Materials (Trace) Subcommittee of the Chemistry-Instrumental SAC

FROM: Lynn Garcia, LRC Liaison to Chemistry-Instrumental SAC

RE: **LEGAL RESOURCE COMMITTEE (LRC) COMMENTS ON E2926-13**

The Materials (Trace) Evidence Subcommittee of the Chemistry-Instrumental SAC is proposing that ASTM E2926–13 (“Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (μ -XRF) Spectrometry”) be placed on the OSAC Registry of Standards and Guidelines.

Our comments are primarily intended to enhance the value of the Standard to the legal community. This Standard will be most helpful if it not only helps assure high quality results in the laboratory, but also is written to show how work performed in accordance with the Standard is both well grounded in theory and data and that it is presented within the boundaries of “the knowledge and experience of [the expert’s] discipline.”¹ Consequently, the comments are intended to address four questions that are important to the legal reception of the Standard:

- (1) Is the Standard written as clearly as possible, and without undefined technical terms and symbols, so as to enable lawyers and judges to grasp the main ideas and requirements set forth?
- (2) Does the Standard describe in detail how the peer-reviewed and readily available scientific literature establishes the validity of the assumptions underlying the scientific tests and the interpretation of test results?
- (3) Does the Standard list the limitations of the tests and results and provide for expressions of the uncertainties in measurements and inferences drawn from them?
- (4) Does the Standard include recommendations or requirements for the creation and retention of documentation of the test and the contents of reports, including the scientific limitations of the tests and related conclusions or inferences?

These are matters of both technical merit and legal importance. Though the LRC is not able to assess the scientific merit of a Standard, our review encompasses whether a Standard makes a

¹ Kumho Tire Co. v. Carmichael, 526 U.S. 137, 148 (1999) (quoting Daubert v. Merrell Dow Pharms, Inc., 509 U.S. 579, 592 (1993)).

prima facie case for the validity of the methods and legal utility of the kinds of expert opinions that a Standard contemplates.

The LRC received feedback from the FSSB recently that it would be more useful for LRC members to provide consolidated comments as opposed to providing the comments of individual members and indicating which other members of the LRC join in the comments. We did not have sufficient time to attempt this before comments on E2926-13 were due. However, we have been discussing possibilities for meeting this request and will strive to make our comments as useful as possible to the FSSB and other interested readers.

Comments by LRC Member David Moran:

1.3. I object to the language "This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment." This language strikes me as (1) unnecessary; and (2) a dodge that a bad forensic scientist could use to justify wholesale deviations from the approved method (i.e., "My professional judgment, experience and training justifies my decision to declare that the glass in this case is not excluded from the reference sample even though the variances exceeded three standard deviations").

This comment is non-persuasive. See general comments provided.

3.5. I don't understand why determining the area under the peaks and comparing that area to the area under peaks of certain elements is considered "semi-quantitative analysis" as opposed to "quantitative analysis."

"Semi-quantitative" refers to comparing ratios of peak areas. Peak areas are related to actual concentrations, which are not calculated by this method. "Quantitative" would refer to determining actual concentrations of the elements, which is not done for u-XRF examinations of glass. See general comments provided.

10.7.3.1 and 10.7.3.2 set out a nice quantitative method allowing the examiner to conclude that two glass specimens are not from the same source. But the standards say nothing about what the examiner should report or say if the method does not result in exclusion. The standard should explicitly say that in that case the examiner should report that the samples cannot be excluded as being from the same source, and nothing more.

This comment is non-persuasive. The suggestion is beyond the scope of the standard. See general comments provided.

Additional Comments:

1. The standard should require the examiner to attach to a written report all of the charts

containing the peaks used to perform the analysis.

This comment is non-persuasive. The suggestion is beyond the scope of the standard.
See general comments provided.

2. The standard should require that the examiner test several "suspect" glass samples, not just one, against the known glass, and that, wherever possible, the examiner should not know which of the "suspect" glass samples is suspected as matching the known glass. In other words, the examiner should be blinded by having someone else choose the samples to be tested without telling the examiner which is the suspected sample.

This comment is non-persuasive. Each Q fragment is necessarily treated as a separate entity, while the K has multiple fragments analyzed in order to more fully characterize the known sample. This suggestion would diminish the usefulness of the comparison.

The following members of the LRC agree with comments made by David Moran:

Barry Scheck and Lynn Garcia join in David Moran's comments. David Kaye joins with the clarification that he would not foreclose the possibility of a revision to the Standard that would allow some scientifically defensible explanation of the implications of a failure to exclude. He also questions whether the statistical decision rule for exclusions based on "peak intensity" is acceptable as written.

Comments by LRC Member David Kaye:

The current ASTM Standard Test Method contains valuable guidance for forensic science laboratories. Nevertheless, I believe that this Standard Test Method can be significantly improved and should not be included in the Registry without substantial changes. If the Standards developed or selected for inclusion on the OSAC registry are intended to represent the kind of "controlling standards" spoken of in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), and thus to be of the most benefit to the law, the Standard should clearly delineate what analysts must do (invariably or in specific circumstances); what they might do; and if current practices are deficient, what they should or must not do. It should indicate the limitations on conclusions that an analyst can reach and should explain and justify its choices (perhaps in a separate supporting document). It should outline the minimum content that good scientific practice dictates for written reports or other laboratory documents. A lay reader should be able to use the Standard to help determine whether a laboratory is providing scientifically reputable testimony in a given case.

Admittedly, this is a tall order, but even if one rejects this aspiration for standards placed in the OSAC registry, perhaps on the theory that a narrowly written Standard can specify enough of what should be done to make it ready for inclusion, the treatment of the matters that the Standard does address should avoid unnecessary ambiguity, should document the validity and reliability of the procedures it prescribes or recommends, and should explicitly state what it does not cover. There is little or nothing to be gained by rushing to endorse Standards that lack these features. Only

Standards that accomplish these goals can fulfill the claim made in the Technical Merit Worksheet for this Standard that it is “fit for purpose” in that “this document can be used as a reference by any law enforcement agency, judges, prosecutors and defense attorneys.”

1. Concerns Regarding Content

Introduction

The introductory paragraph assumes that it is necessary or desirable to reach a binary conclusion (“distinguishable” or “indistinguishable,” which, in court, translates into excluded or included) when comparing two objects. Moreover, it suggests that the “the possibility that they [fragments] originated from the same source of glass” must “be eliminated” for the analysis to be useful. Neither proposition should be endorsed unequivocally.

This comment is non-persuasive. The objective of a forensic glass examination is to compare glass samples to determine if they can be discriminated using the physical, optical, and chemical properties. Ultimately, the goal of this standard test method is to determine if glass samples collected are distinguishable or indistinguishable based on the elemental composition using u-XRF. We disagree with the suggestion that an elimination is the only useful aspect of this analysis. Finding two samples indistinguishable is also useful.

First, neither the introduction nor the rest of the Standard explains why the examiner must use a binary classification as opposed to reporting the probability of the observed degree of similarity if the questioned fragments originated from the known glass versus that if they did not. Second, the “possibility” that different specimens have a common source never can be eliminated. The data can be quite improbable if they originated from the same source. Or, they can be much more probable under the same-source hypothesis than a different-source hypothesis. Using a sharp cutoff for exclusion carries a probability of statistical error. If a cutoff is the only permissible way to interpret the measurements, as the introduction and Part 10 intimate, this uncertainty must be acknowledged.

This comment is non-persuasive. See general comments provided for an explanation of the match criteria.

Part 1

Section 1.3 states that “This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment.” On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc

manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis).

This comment is non-persuasive. This is not the intended meaning of the statement. See general comments provided.

Parts 3-9

Some of the material in Parts 3-9 is descriptive, and some is prescriptive. Statements such as “Limits of detection (LOD) are dependent on several factors, including ...” do not supply much guidance. How should or must limits be set? Steps that are required should be designated as such; those that are merely recommended should be phrased accordingly. There are “musts” and “shoulds” in these sections, but it is not always clear why some of the “shoulds” are not “musts” and what some other things are.

This comment is non-persuasive. LODs were calculated based on inter-laboratory studies following the ASTM guidelines for reporting limits of detection. See general comments section for more detail.

The should/must wording was addressed during the drafting and during the overall balloting process of the ASTM documents. See general comments provided about the ASTM process.

Part 10 (Calculation and Interpretation of Results)

Full sentences (with subjects) or some other wording should be used so it is clear which tasks are mandatory, recommended, or permissible.

This comment is non-persuasive. The should/must wording was addressed during the drafting and during the overall balloting process of the ASTM documents. See general comments provided about the ASTM process.

Section 10.2 could be clearer in stating that automated peak identification and purely manual identification are both acceptable (if they are). Are there other acceptable methods of manual verification of an automated determination?

This comment is non-persuasive. Section 10.2 states that automatic peak identification shall be manually verified by any of the three methods reported in the method.

What use should be made of the “visual comparison”? Can it override quantitative measurements? How should it be performed?

This comment is non-persuasive. Sections 10.6.2 and 10.7.2 address this comment. The visual comparison is a check to see if the samples have obvious elemental differences or to see if semi-quant comparisons are necessary. The semi-quant comparisons are used when there aren't apparent visual differences. There is not a circumstance in which they would be similar in peak intensity ratios but visually spectrally different.

Section 10.7.3.1 states that "If the ranges of one or more elements in the questioned and known specimens do not overlap, it may be concluded that the specimens are not from the same source." The phrase "may be" is rather weak. Is this the recommended conclusion? Why? If the statistical properties of the "ratio ranges" are unknown, how can one know what to conclude? There may as few as 3 measurements of the questioned glass and 9 of the known one.

This comment is non-persuasive. (10.7.3.1): XRF analysis is one of several steps within the glass analysis scheme. There may be other considerations that may prevent an examiner from a definitive elimination in specific cases. For example, if there is a single, very small fragment from a glass container that was recovered from the bottom of a shoe which was very similar to the known glass in all except one ratio, there is the possibility that the known sample was collected in a way that did not provide appropriate characterization of the known sample. Or the possibility exists that a small contaminant particle was not able to be removed or avoided in analysis. The wording in the method allows for considerations such as these.

With respect to the statistical properties of ratio range overlap, the statement that is present acknowledges that the confidence level is not directly addressed. However, research has shown that it is an appropriate method based on the goal of minimizing Type I and Type II errors. See general comments provided about the ASTM requirements for listed references.

Section 10.7.3.2 6 adopts a 3-standard-error rule. It reads

For each elemental ratio, compare the average ratio for the questioned specimen to the average ratio for the known specimens $\pm 3s$. This range corresponds to 99.7 % of a normally distributed population. If, for one or more elements, the average ratio in the questioned specimen does not fall within the average ratio for the known specimens $\pm 3s$, it may be concluded that the samples are not from the same source.

I have puzzled over these sentences for hours without being able to understand them. Is this a decision rule based on a desired 99.7% confidence interval for true mean of the ratio in a homogenous known glass sample? If so, it does not account for the fact that with a standard error estimated from a small sample, one needs a larger interval to achieve 99.7% confidence. In addition, the usual (and better) way to test whether two sample means are different is to use the sampling distribution of the difference between the sample means rather than the sampling distribution of only one of the sample means. Furthermore, even with the proper test statistic and distribution, the many separate tests (one for each ratio Ca/Mg, Ca/Ti, Fe/Zr, etc.) cloud the interpretation of the significance of the difference in a pair of sample means. The risk of a false exclusion for, say, ten comparisons could be ten times the nominal value of 0.003. Thus, the section should be rewritten to justify the choice of the nominal level and to indicate how the nominal level relates to the actual level. In other words, why the 3σ rule? Is it supposed to keep the risk of a false exclusion to a low level?

This comment is non-persuasive. (10.7.3.2): Research has shown that 3s is an appropriate method for elemental comparison of glass by u-XRF based on the goal of minimizing Type I and Type II errors. See general section comments for a detailed explanation on match criteria.

Although these questions may seem technical, they are directly related to the interpretation of the results in the criminal justice system. From a legal perspective, are not false inclusions the type of error that should be guarded against more assiduously? And even if 3σ is the right rule here, why is the standard for making associations via elemental compositions in E2330 some kind of 4σ rule? Without reconciling the different standards, their value as justifications for interpretations of test results in the legal system could be jeopardized.

This comment is non-persuasive. Comment 1: u-XRF is just a part of the overall scheme of glass analysis. Each test assesses properties that are known to be variable among the overall population of glass. Each test is an attempt to find out if there are verifiable and distinguishable differences between two or more samples of glass. At the end, there is typically either an elimination (the Q glass did not originate from the K glass source) or the inability to distinguish the glass samples (the K glass source is a possible source of the glass). In the latter case, there are other possible sources of the Q glass. There is no identification of source using class properties. It is circumstantial evidence which may have explanations other than hypothesis that the Q glass came from the K glass source. The u-XRF testing (or any one test) cannot address comparisons to all other sources in the world, but rather (typically) only to the submitted and tested evidence. As such, it is accurate to say that based on the tests conducted, the source of K glass is a possible source of the Q glass if it can't be distinguished. False inclusions are guarded against,

but class evidence has, by its nature, the possibility of other sources being the actual source. The significance of these other possibilities is beyond the scope of this standard test method.

Comment 2: The match criteria are different because they are different tests and, as such, have different levels of precision and testing protocols. In addition, LA-ICP-MS is a quantitative method, while XRF is not. Research was conducted in order to discover which match criteria would minimize Type I and Type II errors (see general comments section).

Assuming that the exclusion-inclusion decision is the best way to interpret the differences in “peak intensity,” in discussing the interpretation of the data, the Standard needs to offer guidance about the probative value of an inclusion. Should the analyst report that the questioned fragment might have come from the known glass *or from any other glass with a similar set of elemental concentrations*? What data are there on the population distribution of these statistics? If there are none, what can or should the analyst report?

Because u-XRF is only one of a series of tests conducted on glass, and the data from each test is collectively addressed by the examiner, it is inappropriate to provide guidance to report wording for glass that is indistinguishable by XRF. This is outside the scope of the method. Refer to the general comments provided for information regarding statistical methods of interpretation.

Part 11 (Precision and Bias)

These quantities should be defined within the Standard itself.

This comment is non-persuasive. Definitions of particular technical terms are addressed in other ASTM documents. The precision and bias for u-XRF of glass within this standard test method in Section 11.4 and the Appendix.

References

To show a legal audience that the Standard is based on a complete review of the scientific and statistical literature, there should be references to studies that help demonstrate the value of the testing in forensic investigations. The Standard should show how it flows from and is supported by a body of cited scientific studies. The Subcommittee commendably listed 15 papers in the Technical Merit Worksheet of 7/1/15, but readers in the legal community will not know of them.

(Could this be remedied by having an appendix that justifies key choices made in the Standard placed on the registry along with the Standard? This appendix could explain which publications support which choices and how they do so. It would help prevent members of the legal community from misjudging the Standard as the kind of “ipse dixit” condemned in *General Electric v. Joiner*, 522 U.S. 136 (1997). Evaluating this document together with the body of the Standard would permit readers who are not already technical experts in the field to judge its readiness for addition to the registry. Presumably, this explanatory document would not have to be approved by ASTM.)

This comment is non-persuasive. The intended audience of these ASTM documents is the forensic practitioners or scientists that conduct the glass analysis. See general comments section regarding reference/citation policies for ASTM methods.

Concluding Comments

Recommendations about the opinions and conclusions that analysts can reach and how they should present them in reports and in court should be made in light of current thinking about the methods for interpreting and evaluating evidence across the entire domain of forensic science. The premise of the Standard is that the expert’s task is to decide whether a source hypothesis is true or false. Would a likelihood ratio be a better way to express the probative value of the data? Certainly, there is an argument to that effect in the legal and forensic science literatures. See, e.g., Colin Aitken & Franco Taroni, *Statistics and the Evaluation of Evidence for Forensic Science* (2d ed. 2004); James M. Curran et al., *Forensic Interpretation of Glass Evidence* (2000); ENFSI *Guideline for Evaluative Reporting in Forensic Science* (2015); David H. Kaye et al., *The New Wigmore: Expert Evidence* (2d ed. 2011); Royal Statistical Soc’y Working Group on Statistics and the Law, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses* (2010).

These issues have been addressed in the above comments.

In the end, the subcommittee, the SAC and the FSSB may conclude that only the categorical decision framework that is recommended in this document is acceptable for the evaluation and explanation of the evidence in court. If that is their conclusion, however, some of the reasoning behind the conclusion should be provided to assist the legal community in using the measurements wisely and fairly.

2. Drafting Problems

Introduction

Should the last sentence in the introduction read “Measuring elemental concentrations with micro X-ray fluorescence spectrometry permits high discrimination among different sources of glass”?

This comment is non-persuasive. Minor editorial changes will be addressed in the document during the ASTM revision process.

Part 1 (Scope)

I suggest rewording 1.1 to read as follows: “This Standard concerns μ -XRF analysis using mono- and poly- capillary optics, and an energy dispersive X-ray detector (EDS) for the determination of the concentrations of major, minor, and trace elements in glass fragments.” This change would permit 1.2 to be deleted.

This comment is non-persuasive. Minor editorial changes will be addressed in the document during the ASTM revision process.

1.3 states that “This test method does not replace knowledge, skill, ability, experience, education, or training and should be used in conjunction with professional judgment.” On its face, this seems to assert that analysts can ascertain elemental composition without using any instruments or that an analyst can depart from one of the prescribed statistical rules in an ad hoc manner. The sentence should be clarified (or deleted on the theory that it goes without saying that it takes skill, experience, and judgment to perform the analysis).

This is not the intended meaning of the statement. See general comments provided.

1.4 reads: “The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.” The meaning of “to be regarded as standard” is not immediately apparent. Why is it necessary to say that no units other than the International System of Units are used? Is not that apparent from reading the Standard as a whole? (Of course, if a Standard would not be expected to use these units, then referring to this choice at the outset is helpful.) Moreover, this remark does not really describe the scope of the Standard, but only how things within its scope are expressed.
See general comments section.

The disclaimer in section 1.5 does not seem to achieve any legal objective. It is hard to see how any reader would think that a “method . . . purport[s] to address all of the safety concerns,” and stylistically, the phrase “all of the safety concerns, if any” also is awkward, and the “if any” phrase is contradicted by the fact that Section 7 does address a safety concern.

Sections 1.4 and 1.5 are required by ASTM methods.

Part 10 (Calculation and Interpretation of Results)

What does it mean to “correct” sum peaks and escape peaks? That the analyst should label them as such?

Correcting sum and escape peaks is recognizing and labeling them, as some software programs may assign the peaks to a different element.

The following members of the LRC agree with the comments made by David Kaye:

Barry Scheck and David Moran join in David Kaye's comments.

Ron Reinstein joins in David Kaye's comments except the comment regarding Section 1.3. Judge Reinstein believes this section should be clarified but not deleted. Judgment, training and experience are important and must be used in conjunction with the test method (but not in place of it).

These issues have been addressed in the above comments.

Judge Reinstein and Lynn Garcia would like to emphasize the importance of David Kaye's "Concluding Comments" on page 6—this is the type of comment that the SACs and FSSB should pay particular attention to for standards that are to be included on the OSAC Registry.

Additional Comments by LRC Member Barry Scheck:

With respect to David Moran's comments, Barry Scheck would like to emphasize the term "semi-quantitative" is troubling. Either you have data or you don't and the measure uncertainty can be calculated within acceptable limits. If you are going to rely on "experience" to declare exclusions or inclusions (which seems problematical in the first place) the standard should specify when and how that would be done and any validation that justifies it.

This comment is non-persuasive. "Semi-quantitative" is not implying partial data. The term is used because the concentrations of elements are not being quantified, but rather, the peak areas (which correspond to concentrations even when those concentrations aren't explicitly known) are used in ratios to one another. See general comments section for more details.

Barry Scheck would also like to emphasize concerns expressed in the comments by David Kaye that the deficiencies in the statistical explanations are troubling and not ready for court, whether one is in a *Frye* or *Daubert* jurisdiction. These should be rejected from the OSAC Registry and, hopefully, the OSAC subcommittee and/or ASTM will revise the proposed standards to follow the template laid out in the Technical Merit Worksheets. The requirement of general acceptance in the scientific community, particularly among statisticians, cannot be met, nor the requirements of clearly identifying limitations and weaknesses in the methodology or an explanation of how it is "fit for purpose."

These issues have been addressed in the above comments.

DISCLAIMER: *The failure of any member of the Legal Resource committee (LRC) to provide a comment, identify a legal issue or join in another LRC comment should not be interpreted as a disagreement or endorsement of the comment, the standard or its legal sufficiency.*