

To: OSAC Independent Review Panel

From: David Kaye, Ronald Reinstein, and Barry Scheck

Subject: Appeal from Adjudication of LRC-compiled Comments on ASTM E2926-13

Date: April 3, 2017

## Introduction

The FSSB approved ASTM E2926–13, a “Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence ( $\mu$ -XRF) Spectrometry” over objections from the FSSB Statisticians Task Group and the Legal Resource Committee (LRC). This appeal from members of the LRC is required because the adjudication process failed to address significant parts of certain comments that would have helped the subcommittee produce an improved document for the benefit of the forensic science and legal communities.

## Appeal

### I. Comment on § 1.3

The LRC-compiled comment on § 1.3 is that

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).<sup>1</sup>

The response was

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

Obviously, labeling a comment as “non-persuasive” is not a sufficient response to a request for a change. There must be a reason given to believe the comment does not merit a change. The comment sought a standard that would clearly express the intent behind the sentence. Rather than consider how clear the proposed sentence is and whether a change would improve it, the subcommittee asserted that it did not intend for it to be read in the way that the

---

<sup>1</sup> Judge Reinstein did not agree with the parenthetical part of the comment. Consistent with the overall comment, he proposed a simple change to the wording to avoid ambiguity—that “judgment, training, and experience are important and must be used in conjunction with the test method, but not in place of it.” He continues to believe that a sentence to this effect belongs in the standard.

comment pointed out it might be read.<sup>2</sup> When confronted with ambiguous language, however, the proponents of the text cannot just say that they did not intend it to be ambiguous. They could reply that it is not really ambiguous (and explain why) or make a change to eliminate the ambiguity. The response here does neither. It does not engage the issue at all.

## II, Comment on Missing Words in Part 10 (Calculation and Interpretation of Results)

The following comment was not adjudicated properly:

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

The comment simply asked the subcommittee to identify which steps in the procedure as enumerated are required and which are not.

The response was *not* that Part 10 already does this well enough; nor was it that more clarity would be counter-productive. Instead, the putative answer was

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.

The general comments state that

As per the ASTM internal guide, documents require language consistency: use the word shall when stating mandatory requirements, use the word should as advisory, use the word may to indicate optional directives, avoid use of must whenever possible. The whole process of changing shall/should/must requires new balloting. Both of the revised methods were developed keeping this in mind with thorough consideration of the practical implications on when/why to use one term over the other.

We are at a loss to understand how the fact that ASTM requires “shall” for a requirement, “should” for a recommendation, and “may” for neither, and that it eschews “must”—a word that, inexplicably, appears nine times in this ASTM standard—responds to the comment that Section 10 often omits these words—a drafting flaw that causes ambiguity.<sup>3</sup> Nor do we understand how the fact that an SDO formulated a standard with particular wording can be considered a response to a suggestion for improving the wording. If this were an admissible response, all subcommittees could simply say that they are putting forward an SDO-approved standard for the registry and need not reply further to comments on the merits of those standards.

## III. Comment on § 10.7.3.2

Section 10.7.3.2 6 reads:

---

<sup>2</sup> The general comments merely state the *intended* meaning and explain that the reason for the words selected was that they are “highly recommended by the E30 ASTM committee.” The general comments do not explain why or how the “highly recommended” wording found by members of the LRC to be problematic conveys the subcommittee’s intended meaning.

<sup>3</sup> Section 10.1, for example, consists of the sentence fragment “Examine the spectrum, and identify and label the peaks.” Is this directive a “shall,” a “should” or a “may”?

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.

The LRC-compiled comment, in pertinent part, is as follows:

Is this a decision rule based on a desired 99.7% confidence interval for the 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.

It is not necessary to understand all the technicalities of the comment (and the subcommittee's general comments) to see that the bottom line is a request for the standard to *clarify* the fact that the even though 99.7% of the area under a normal curve lies within approximately plus-or-minus three standard deviations of its mean, "compar[ing] the average ratio for the questioned specimen to the average ratio for the known specimens  $\pm 3s$ "<sup>4</sup> across many elements *will not necessarily achieve 99.7% confidence*.

The adjudication rejects recognizing this issue in the standard, not on the ground that the matching rule will produce 99.7% confidence as promised or suggested, but rather because studies not mentioned in the standard prove that the actual confidence is "appropriate" (even if it is different than the figure in the standard):

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

This appeal does not question the judgment that the empirically established conditional error rates are "appropriate."<sup>5</sup> That part of the adjudication (and every other part of it) can be taken as true for the purpose of this appeal. We only contend that nothing in the adjudication of

---

<sup>4</sup> The subsection does not define "s," but the symbol often refers to a sample standard deviation as distinct from the true but unknown standard deviation ( $\sigma$ ) of the sampling statistic. The 99.7% figure applies to a confidence interval of approximately  $\pm 3\sigma$ .

<sup>5</sup> The FSSB's Statistics Task Group disputed this conclusion. Although Appellant Kaye is a member of that group, he did not participate in the STG vote on this question and took no position on the adequacy of the studies said to establish that the conditional error probabilities of the  $\pm 3s$  rule are appropriately small.

comments actually responds to the request for a clarifying statement in the standard.<sup>6</sup> The adjudication neither adheres to the figure of 99.7%<sup>7</sup> nor explains why the sentence should not be modified (other than, perhaps, a general suggestion that it would be inconvenient or time-consuming to require re-balloting within ASTM).

#### **IV. Comments Seeking Editorial Changes**

A few LRC-compiled comments seeking to improve wording (but that were not needed to resolve ambiguities or to correct assertions) were met solely with the reply that

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

A promise of future change by a different organization is not an acceptable basis for an adjudication of comments on a proposal to approve a draft in its current form. (This principle was accepted in an appeal from the adjudication of ASTM E2548-13.) The OSAC adjudication process asks for editorial as well as substantive comments. If subcommittees are unwilling to consider minor editorial changes, the process should be revised to reflect this limitation on the process for potentially revising documents before placing them on the OSAC registry of approved standards.<sup>8</sup>

\* \* \*

For the reasons stated above, the adjudication process with respect to the three sets of LRC-compiled comments listed here was flawed.

---

<sup>6</sup> Judge Reinstein notes that the statistical content of section 10.7.3.2 is beyond his expertise.

<sup>7</sup> The Statistics Task Group confirmed that the standard's suggestion that the theory that a normal distribution of measurement error implies 99.7% confidence for the  $\pm 3s$  rule (as applied to six or seven elements) is invalid. Although it goes beyond the scope of this appeal, we believe that the adjudication process should include input from a Statistics Resource Committee so that when subcommittees adjudicate comments, they will be fully informed with respect to the statistical aspects of their standards. The Legal Resource Committee urged this enhancement at the OSAC Leadership Strategy Session held on June 22, 2016.

<sup>8</sup> The purely editorial changes proposed for this standard were minor. If no other procedural errors in the adjudication had occurred, the appeal panel might have been able to treat them as "harmless error" not justifying further adjudication and FSSB review.