LABORATORY PERFORMANCE CRITERIA IN THE ENVIRONMENTAL LEAD PROFICIENCY ANALYTICAL TESTING (ELPAT) PROGRAM

Paul C. Schlecht
Division of Physical Sciences and Engineering

Ruiguang Song, Ph.D.
Computer Sciences Corporation/DPSE, NIOSH

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health
Division of Physical Sciences and Engineering
Robert A. Taft Laboratories
4676 Columbia Parkway
Cincinnati, Ohio 45226-1998

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

December 1996

NIOSH

Laboratory Performance Criteria in the Environmental Lead Proficiency Analytical Testing (ELPAT) Program

Paul C. Schlecht[1] and Ruiguang Song[2]

[1] Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226, USA

[2] Computer Sciences Corporation/NIOSH, 4676 Columbia Parkway, Cincinnati, Ohio 45226, USA

ABSTRACT

This article examines the rating criteria used in a laboratory proficiency testing program that is based upon a laboratory's performance over one year. Statistical power of the current criteria is calculated. Alternative criteria based on z-scores are proposed. Compared to the current criteria, the proposed criteria are more powerful in detecting poor laboratory performance and more flexible in detecting bias or poor precision. The proposed criteria can be easily adjusted for various needs and the estimates of bias and precision of each participating laboratory can be obtained using the z-scores.

Key Words: Accuracy, Bias, ELPAT, PAT, Precision, Rating Criteria, Statistical Power, Z-Score.

Mr. Schlecht is Chief of the Quality Assurance and Statistics Activity. He is a member of ASQC.

Dr. Song is a mathematical statistician in the Quality Assurance and Statistics Activity.

DISCLAIMER

Introduction

To ensure the quality of analyses associated with lead abatement, a proficiency testing program, the Environmental Lead Proficiency Analytical Testing (ELPAT) Program, was started in the Fall of 1992 (see Groff and Schlecht, 1996b, or Schlecht and Groff, 1996). The ELPAT Program rates laboratories based upon their analysis of lead containing samples. Over the first three years of the ELPAT Program 88-92% of participating laboratories are rated proficient. These ELPAT ratings are then used in conjunction with onsite evaluations of laboratory operations by two organizations, the American Association for Laboratory Accreditation (A2LA) and the American Industrial Hygiene Association (AIHA) to accredit laboratories. A2LA and AIHA accreditations are then used by the U.S. Environmental Protection Agency (EPA), the National Institute for Occupational Safety and Health (NIOSH), and the National Safety Council (NSC) to recommend qualified laboratories to the public. As of early 1996, over 400 environmental lead laboratories participating in the ELPAT Program. Results of the program are published in quarterly columns of the American Industrial Hygiene Association Journal (AIHAJ) and American Conference of Governmental Industrial Hygienists (ACGIH) Applied Occupational and Environmental Hygiene Journal (AOEHJ).

The ELPAT Program uses a simple criteria based upon the number of acceptable results to rate a laboratory's performance (Esche, Groff, Schlecht, and Shulman, 1994). In this article, the statistical power of the current criteria is calculated. Alternative criteria based upon z-scores and a harmonized proficiency testing protocol (Thompson and Wood, 1993) have been proposed for the ELPAT Program. With the z-scores, bias and precision of a participating laboratory can be estimated. Then, an accuracy score depending on the bias and precision is used as the test statistic. Tables showing the statistical power for the alternative criteria are provided. Compared to the current criteria, the proposed criteria are more powerful in detecting poor laboratory performance and more flexible in control of detecting bias or detecting poor precision and can be easily adjusted for various needs.

The ELPAT Program and Its Rating Criteria

Samples of three sample types - paint chips, soils, and dusts on wipes at four lead levels for each sample type are sent to participating laboratories each quarter (termed a round). Laboratories are evaluated at NIOSH by comparing each laboratory's result against an acceptable performance range. The acceptable performance range is based upon consensus values from reference laboratories and is modeled after the evaluation procedures currently used in an industrial hygiene proficiency testing program, the Proficiency Analytical Testing (PAT) Program (see Schlecht and Groff, 1995, or Groff and Schlecht, 1996a). The criteria proposed in this article could be applied to rate the over 1400 industrial hygiene laboratories participating in the PAT Program.

Laboratories which have been rated proficient on the previous ELPAT round and are accredited by either A2LA or AIHA are selected as reference laboratories. Results from reference laboratories are used to determine the performance limits for each sample. After data from reference laboratories are collected and extreme reference laboratory results are Winsorized to the highest result (for the top 5 percent) or the lowest result (for the bottom 5 percent) remining in the set, the mean ± 3 standard deviations of the reference laboratory data become the acceptable performance range. (Using a large number of laboratories (>30) as reference laboratories and treating reference laboratory outliers reduce the influence that any one reference laboratory has on ELPAT performance limits.) Laboratory results are acceptable if they fall within the performance limits. Results falling outside the performance limits are designated as outliers. Laboratories are rated based upon their performance in the

ELPAT Program over the last year (i.e., four rounds) for each lead sample type. A laboratory is proficient for the lead sample type if:

1. All four results on each round have been reported and all eight results are designated as acceptable for the last two consecutive rounds; or
2. Three-fourths or more of the results reported in the last four consecutive rounds (up to 16 results) are designated as acceptable.

A rating of nonproficiency indicates that a serious analytical problem has been identified and warrants immediate corrective action by the laboratory and follow-up by cooperating laboratory accreditation organizations. A nonproficient rating also results in immediate removal of the laboratory from the list of laboratories recognized by the EPA National Lead Laboratory Accreditation Program (NLLAP) which is used by EPA, NIOSH, and the National Safety Council Hotline to recommend qualified laboratories to the public.

Power Calculation Based on the Current Criteria

A proficiency rating in the ELPAT Program is actually a statistical hypothesis test. Each participating laboratory is compared to reference laboratories. A laboratory is rated proficient if its results agree with ELPAT reference laboratory results. Under the assumptions given below, a laboratory's performance can be characterized by two parameters: bias and precision. The bias is the relative deviation from the true value and the precision is the uncertainty measured by relative standard deviation. Since the true value is unknown, the bias is defined as the relative deviation of a laboratory's mean from the reference laboratory's mean and the reference laboratories are assumed to be unbiased.

To evaluate the current rating criterion and later to compare it with the proposed rating criterion, we need to examine the statistical power of the testing program under each criterion. To compute the power, assumptions are needed although they are not required for implementing a rating criterion. The assumptions that are reasonable and almost satisfied in the ELPAT Program are:

(1) All measurements are independent from laboratory to laboratory, round to round, and sample to sample.
(2) All measurements are normally distributed random variables.
(3) All reference laboratories have an identical distribution at each concentration level with a mean equal to the true value, and have a constant precision for all concentrations and for all testing rounds.
(4) Each laboratory under test has a constant bias and a constant precision for all concentrations, and for all testing rounds.

To define the bias and precision clearly, let T be the true value and x be a measurement of T by the laboratory under test. Suppose that the mean and the standard deviation of the random variable x are μ and σ_t, respectively. Then, the bias and precision of this laboratory are defined by

$$B = (\mu - T)/T \quad \text{and} \quad CV_t = \sigma_t/\mu,$$

respectively. Here, bias is a relative bias and precision is a relative standard deviation (RSD) or coefficient of variation (CV). Both the bias B and the precision CV_t are assumed to be constants. For a reference laboratory, the measurements are assumed to be unbiased (mean = T) and the precision, denoted by $CV_r = \sigma_r / T$, is also a constant. Based on these assumptions, the ELPAT Program is testing participating laboratories with hypotheses

$$H_0\text{: } B = 0 \text{ and } CV_t = CV_r \quad \text{vs} \quad H_1\text{: } B \neq 0 \text{ or } CV_t > Cv_r.$$

The null hypothesis implys that the participating laboratory does not differ from the reference laboratories in performance and the alternative hypothesis implys that the participating laboratory differs from the reference laboratories with either different mean or worse precision.

Let X and Y be the numbers of results not acceptable in the first two rounds and the last two rounds, respectively. Under the assumptions above, X and Y are mutually independent, and both of them have the same binomial distribution B(8,q), where q is the probability of having an outlier (a result not acceptable) in one analysis. According to the current ELPAT criteria, a result is an outlier if it is outside the performance limits, which are the sample mean ± 3 standard deviations of reference laboratories. This probability can be calculated by

$$q = Prob(\ |x-\bar{x}_r| > 3\hat{\sigma}_r\ )$$

where $\bar{x}_r$ and $\hat{\sigma}_r$ are the sample mean and sample standard deviation of reference laboratories' results.

Let

$$z = (x-\bar{x}_r)/\hat{\sigma}_r \tag{1}$$

and

$$\lambda = \sqrt{\sigma_t^2/\sigma_r^2+1/n}.$$

Then under our assumptions, z/λ has a noncentral t-distribution with n-1 degrees of freedom and a noncentrality parameter $\delta = B/(\lambda CV_r)$. Define $\rho = \sigma_t / \sigma_r$, a ratio of the two standard deviations (STD). Then the STD ratio $\rho = (B+1)CV_t/CV_r$ is a constant by our assumptions. Therefore,

$$q = Prob(\ |t_{n-1}(\delta)| > 3/\lambda\ ). \qquad (2)$$

Let E_2 and E_4 be the event that a participating laboratory is rated proficient by the two- and four-round criteria and $\bar{E}_4$ be the complement event of E_4. Note that X and Y are mutually independent and X + Y has a binomial distribution B(q,16). Given the relative bias B and the ratio ρ, the probability that a participating laboratory is rated proficient by the four-round criterion is

$$p_4 = Prob(E_4) = Prob(X+Y \leq 4) = \sum_{i=0}^{4} C_{16}^{i} q^{i} (1-q)^{16-i},$$

where $C_m^k = m!/(k!(m-k)!)$ is the binomial coefficient.

The probability that a participating laboratory is rated proficient by the two-round criterion is

$$p_2 = Prob(E_2) = Prob(Y=0) = (1-q)^8.$$

The probability that a participating laboratory is rated proficient either by the two-round criterion or by the four-round criterion is then given by

$$
\begin{aligned}
p &= Prob(E_4 \text{ or } E_2) \\
&= Prob(E_4)+Prob(\bar{E}_4 \text{ and } E_2) \\
&= p_4+Prob(X>4)Prob(Y=0) \\
&= p_4+p_2\sum_{i=5}^{8} C_8^i q^i(1-q)^{8-i}.
\end{aligned}
$$

It can be seen that the probability p_2, p_4, and p are functions of q. From (2), the probability q is determined by bias B, STD ratio ρ, and the precision of reference laboratories CV_r. Given ρ and CV_r, the values of q are symmetric about B since the noncentral t-distribution is symmetric about its noncentrality parameter and the noncentrality parameter is symmetric about the bias.

With these formulas, the statistical power (1-p) to detect poor laboratory performance can be calculated. For bias from 0 to 0.3 by 0.05 and the STD ratio from 1 to 3 by 0.5, the probability q and the probability of non-proficient rating $1-p_2$, $1-p_4$, and 1-p are provided in Table 1. In our calculations, we set $n=35$ and $CV_r=0.1$ since these values are typical of the number of reference laboratories and the relative standard deviations of reference laboratories that have been reported over the first two to three years of ELPAT Program operations.

From Table 1, we can see that the chance to fail the two-round criterion is much greater than the chance to fail the four-round criterion. A laboratory has to perform extremely well to pass the two-round criterion. This criterion is designed to provide an opportunity for those laboratories that have poor performance in the previous two rounds to demonstrate their proficiency. Adding the two-round criterion to the program reflects the policy that the most recent performance is more valuable in rating participating laboratories. The two-round criterion does not benefit the laboratories with no improvement in their performance. This can be seen from Table 1 that there are not much difference between the power for the four-round criterion and the overall criterion.

The z-Score and Accuracy Criterion

A statistical rating procedure may have several characteristics. One desirable characteristic is simplicity: that a procedure be easy to implement and understand. Another important feature is the statistical power of the criteria to detect laboratories with poor analysis quality. Gaining some power may require sacrificing some simplicity. In this section, we examine the z-score, which has been suggested in an IUPAC/ISO/AOAC Harmonized Proficiency Testing protocol, adopted by some proficiency testing programs in the United States and abroad (Thompson and Wood, 1993). Considerable effort is underway to develop protocols for proficiency testing which might be used by multiple proficiency testing programs to facilitate international recognition of proficiency testing programs and related laboratory accreditation programs. Based on z-scores, a laboratory's bias and precision can be estimated. To take both the bias and precision into account, an accuracy score derived from NIOSH accuracy criterion is used as the test statistic. (The NIOSH accuracy criterion is a criterion that has been widely used for twenty years in the industrial hygiene field to evaluate analytical methods.)

The z-score defined by equation (1) is indirectly used by the current ELPAT rating procedure by converting the z-score to a binary score as an outlier indicator. By conversion to a binary score, considerable information contained in the z-score is lost. According to the previous analysis, the z-score divided by λ has a noncentral t-distribution. It is known that a t-distribution approaches a normal distribution when its degrees of freedom go to infinity (i.e., Johnson and Kotz, 1970). So, a t-distribution with large degrees of freedom, say greater than 30, can be approximated by a normal distribution with the same mean and the same standard deviation. Suppose that n is the number of reference labs. Then the mean and standard deviation of each z-score can be derived through the mean and standard deviation of the noncentral t-distribution (see Johnson and Kotz, 1970, pages 203-204). They are given by

$$\mu_z = \sqrt{(n-1)/2}\frac{\Gamma(n/2-1)}{\Gamma((n-1)/2)}\delta\lambda \tag{3}$$

and

$$\sigma_z = \sqrt{\frac{n-1}{n-3}(1+\delta^2)\lambda^2-\mu_z^2}, \tag{4}$$

repectively.

Now, let $z_1,\ldots,z_m$ be the z-scores corresponding to results from the laboratory under test. Under our assumptions, these z-scores are mutually independent with an identical distribution. Let $\bar{z}$ and $\hat{\sigma}_z$ be the sample mean and sample standard deviation of the m z-scores. Set

$$\bar{z} = \mu_z \ , \quad \hat{\sigma}_z = \sigma_z \ .$$

Replace CV_r by its estimate and solve the two equations to get estimates for bias B and precision CV_t:

$$\hat{B} = \frac{\bar{z}\hat{CV}_r}{\sqrt{(n-1)/2}}\frac{\Gamma((n-1)/2)}{\Gamma(n/2-1)} \tag{5}$$

$$\hat{CV}_t = \frac{\hat{CV}_r}{\hat{B}+1}\sqrt{(\bar{z}^2+\hat{\sigma}_z^2)\frac{n-3}{n-1}-\frac{\hat{B}^2}{\hat{CV}_r^2}-\frac{1}{n}}\ . \tag{6}$$

For large n, (3) and (4) can be approximated by

$$\mu_z \approx \delta\lambda\ ,\ \sigma_z \approx \lambda \approx \rho.$$

Then, the bias and precision estimates can be simplified as

$$\hat{B} \approx \bar{z}\hat{CV}_r \tag{7}$$

$$\hat{CV}_t \approx \hat{\sigma}_z\hat{CV}_r\ /(\hat{B}+1). \tag{8}$$

With the bias estimate and precision estimate available, we may consider the NIOSH accuracy criterion (Bartley and Fischbach, 1993; Fischbach, Song, and Shulman, 1996). The NIOSH accuracy, denoted by A, is defined as the maximum absolute error relative to the amount T being measured, such that the interval ((1-A)T, (1+A)T) will cover a future measurement x with 95% confidence. That is,

$$Prob((1-A)T<x<(1+A)T) = 0.95$$
$$or \tag{9}$$
$$Prob(-A<\frac{x-T}{T}<A) = 0.95.$$

Under the assumption that x ~ $N(\mu,\sigma^2)$,

$$\Phi\left(\frac{A-B}{(B+1)CV_t}\right) - \Phi\left(\frac{-A-B}{(B+1)CV_t}\right) = 0.95\ , \tag{10}$$

where Φ( ) is the cumulative distribution function of standard normal distribution.

From equation (10), accuracy A is a function of bias B and precision CV_t. According to the NIOSH accuracy criterion, an analytical method should demonstrate its accuracy A $\leq$ 0.25. We may adopt this criterion by estimating accuracy A with bias estimate and precision estimate given by (5) and (6), or by (7) and (8). However, this might be too complicated since there is no explicit expression for the accuracy function. To keep it simple, we may consider using the sample mean and sample standard deviation of z-scores instead of using the bias estimate and precision estimate. Further, we may define a score such that it is a simple function of the mean and standard deviation of the z-score. A simple and general function which takes both bias and precision into account is

$$A_{\alpha,\beta} = a|\mu_z|^\alpha + b\sigma_z^\beta, \tag{11}$$

where a, b, α, and β are fixed for a specific criterion. These constants can be adjusted for the needs of various criteria. The two parameters μ_z and σ_z are estimated by the sample mean and sample standard deviation of z-scores. Such defined accuracy score (11) does not have the meaning defined by (9), but it should be close for some selected values of a, b, α, and β , and by being defined this way, it gives us the flexibility in balancing the two components.

To determine a reasonable critical value, we need to know or be able to calculate the distribution function of the accuracy score estimator. Given a critical value C, the probability Prob(A $\leq$ C) is the probability that a laboratory is rated proficient under the accuracy criterion. This probability can be calculated by

$$Prob(A \le C) = \iint_{a|x|^\alpha + by^\beta \le C} f_{\bar{z}}(x) f_{\hat{\sigma}_z}(y)\,dx\,dy$$

$$= \int_{-\sqrt[\alpha]{C/a}}^{\sqrt[\alpha]{C/a}} \int_0^{\sqrt[\beta]{(C-a|x|^\alpha)/b}} f_{\bar{z}}(x) f_{\hat{\sigma}_z}(y)\,dx\,dy$$

$$= \int_{-\sqrt[\alpha]{C/a}}^{\sqrt[\alpha]{C/a}} f_{\bar{z}}(x) F_{\hat{\sigma}_z}\left(\sqrt[\beta]{(C-a|x|^\alpha)/b}\right) dx$$

$$= \int_0^{\sqrt[\alpha]{C/a}} (f_{\bar{z}}(x) + f_{\bar{z}}(-x)) F_{\hat{\sigma}_z}\left(\sqrt[\beta]{(C-a|x|^\alpha)/b}\right) dx \ .$$

here and afterwards, $f_\xi(x)$ and $F_\xi(x)$ denote the probability density function and cumulative probability function for the random variable ξ, respectively.

Note that the z-score has an approximate normal distribution. So approximately,

$$\bar{z} \sim N(\mu_z, \sigma_z^2/m) \ , \quad \hat{\sigma}_z \sim \sigma_z \sqrt{\chi^2_{m-1}/(m-1)} \ .$$

Therefore, the probability distribution function of the sample mean of z-scores is given by

$$f_{\bar{z}}(x) \approx \frac{\sqrt{m}}{\sigma_z} \phi\left(\sqrt{m}(x-\mu_z)/\sigma_z\right) ,$$

where $\phi(\ )$ is the probability distribution function of standard normal distribution, and

$$F_{\hat{\sigma}_z}(y) \approx F_{\chi^2_{m-1}}\left((m-1)y^2/\sigma_z^2\right) .$$

Set m = 8 for a two-round criterion and m = 16 for a four-round criterion.

Under the conditions that $CV_r = 0.1$ and $n = 35$, the probability for a laboratory getting a nonproficient rating will be a function of parameters B, ρ, a, b, α, β, and C . This probability is symmetric about B because the integrant, especially $f_{\bar{z}}(x)+f_{\bar{z}}(-x)$, is symmetric about B. This can be seen by the following facts:

(1) The sign of B determines the sign of δ .

(2) The sign of δ determines the sign of μ_z .

(3) ϕ is a symmetric function.

The probability values of Prob(A≤C), called the statistical power, for the two-round and the four-round criteria are calculated and listed in Tables 2 and 3, respectively. In each of the two Tables, criteria based on $A_{1,1}$ and $A_{2,2}$ scores with $a = b = 1$ are compared to the current criterion. To achieve similar power with the current criteria, three critical values are selected for each of the two A-scores. From the two Tables, we see the proposed criteria raise power more rapidly than the current criteria when the bias and precision ratio are moderate. The power curves of the proposed criteria are flatter than the current criteria when both the bias and precision ratio are small, or when either the bias or the precision ratio is large. These properties of the proposed criteria are desired because a proficiency testing program should be able to catch the laboratories with poor performance (large bias or poor precision), and meanwhile, let the laboratories with small bias and good precision pass the test. To visualize the power curves and the difference between the current and the proposed criteria, the probabilities of getting a nonproficient rating in four-round by the two criteria are plotted in figures 1 and 2, respectively.

To see the effect of weight (a,b) in the A-score, power values for the four-round $A_{1,1}$-scores with weight (a,b) = (2/3,4/3), (1,1), and (4/3,2/3) are calculated and listed in Table 4. We see that changing the weight results in changing the power. More weight on bias (a>b) makes the criterion more sensitive to the bias and less sensitive to the precision, in other words, more powerful to detect large bias and not so powerful to detect poor precision (see figure 3). Reversing the weight will reverse the sensitivities (see figure 4). So, adjusting the weight can obtain criteria for specific programs with different needs. The difference in power sensitivity can be well seen from the two plots

Summary

In this article, we examined the current ELPAT rating criteria by computing statistical power. The current criteria are simple, but less powerful compared to the proposed criteria that use z-scores. This is because the current criteria do not use all of the information contained in the z-scores. By changing weights of the bias and precision components, the proposed criterion can be adjusted for greater sensitivity to a laboratory's bias or precision. Using the z-scores, one can estimate the participating laboratories' bias and precision.

References

Bartley, D. L. and Fischbach, T. F. (1993). "Alternative Approaches for Analyzing Sampling and Analytical Methods". *Journal of Applied Occupational and Environmental Hygiene* 8, pp. 381-385.

Esche, C. A.; Groff, J. H.; Schlecht, P. C.; and Shulman, S. A. (1994). "Laboratory Evaluations and Performance Reports for the Proficiency Analytical Testing PAT and Environmental Lead Proficiency Analytical Testing ELPAT Programs". DHHS (NIOSH) Technical Report No. 95-104.

Johnson, N. L. and Kotz, S. (1970). *Continuous univariate distributions-2*. Houghton Mifflin Company, Boston, MA.

Groff, J. H. and Schlecht, P. C. (1996a). "PAT Program Report". *American Industrial Hygiene Association Journal* 57(1), pp. 79-81.

Groff, J. H. and Schlecht, P. C. (1996b). "ELPAT Program Report: Background and Current Status(October 1995)". *American Industrial Hygiene Association Journal* 57(2), pp. 208-215.

Schlecht, P. C. and Groff, J. H. (1996). "ELPAT Program Report, Background and Current Status". *Journal of Applied Occupational and Environmental Hygiene* 11(1), pp. 10-16.

Schlecht, P. C. and Groff, J. H. (1995). "PAT Program Report, Background and Current Status". *Journal of Applied Occupational and Environmental Hygiene* 10(12), pp. 967-968.

Thomas J. Fischbach, Ruiguang Song, and Stanley A. Shulman (1996). Some Statistical Procedures for Analytical Method Accuracy Tests and Estimation. To appear in the *American Industrial Hygiene Association Journal.*

Thompson, M. and Wood, R. (1993). "International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories". *Journal of AOAC International* 76, pp. 926-940.

Table 1. The Probability q and Probabilities of Non-Proficient Rating Under the Current (Two-Round, Four-Round, and Overall) Rating Criteria

ρ	$\|B\|$	q	$1-p_2$	$1-p_4$	$1-p$
1.0	0.00	0.006	0.04	0.00	0.00
1.0	0.05	0.011	0.09	0.00	0.00
1.0	0.10	0.033	0.24	0.00	0.00
1.0	0.15	0.085	0.51	0.01	0.01
1.0	0.20	0.182	0.80	0.15	0.15
1.0	0.25	0.329	0.96	0.65	0.64
1.0	0.30	0.508	1.00	0.97	0.97
1.5	0.00	0.055	0.36	0.00	0.00
1.5	0.05	0.068	0.43	0.00	0.00
1.5	0.10	0.106	0.59	0.02	0.02
1.5	0.15	0.172	0.78	0.13	0.13
1.5	0.20	0.265	0.91	0.42	0.42
1.5	0.25	0.379	0.98	0.79	0.78
1.5	0.30	0.506	1.00	0.97	0.96
2.0	0.00	0.144	0.71	0.07	0.07
2.0	0.05	0.156	0.74	0.09	0.09
2.0	0.10	0.192	0.82	0.18	0.18
2.0	0.15	0.248	0.90	0.36	0.36
2.0	0.20	0.323	0.96	0.63	0.62
2.0	0.25	0.411	0.99	0.85	0.85
2.0	0.30	0.506	1.00	0.97	0.96
2.5	0.00	0.239	0.89	0.33	0.33
2.5	0.05	0.249	0.90	0.36	0.36
2.5	0.10	0.275	0.92	0.46	0.46
2.5	0.15	0.318	0.95	0.61	0.61
2.5	0.20	0.374	0.98	0.78	0.77
2.5	0.25	0.440	0.99	0.90	0.90
2.5	0.30	0.513	1.00	0.97	0.97
3.0	0.00	0.325	0.96	0.63	0.63
3.0	0.05	0.332	0.96	0.66	0.65
3.0	0.10	0.351	0.97	0.71	0.71
3.0	0.15	0.382	0.98	0.79	0.79
3.0	0.20	0.423	0.99	0.88	0.87
3.0	0.25	0.472	0.99	0.94	0.94
3.0	0.30	0.527	1.00	0.98	0.98

Table 2. Probabilities of Non-Proficient Rating Under Two Different A-Score Criteria (Compared to the Current Two-Round Criterion)

ρ	$\lvert B \rvert$	$1-p_2$	$A_{1,1} = \lvert M_z \rvert + S_z$ C=2	C=2.5	C=3	$A_{2,2} = M_z^2 + S_z^2$ C=4	C=4.5	C=5
1.0	0.00	0.04	0.11	0.01	0.00	0.01	0.00	0.00
1.0	0.05	0.09	0.31	0.08	0.01	0.04	0.02	0.01
1.0	0.10	0.24	0.73	0.38	0.12	0.22	0.14	0.08
1.0	0.15	0.51	0.95	0.79	0.47	0.63	0.52	0.42
1.0	0.20	0.80	1.00	0.97	0.84	0.92	0.88	0.82
1.0	0.25	0.96	1.00	1.00	0.97	0.99	0.98	0.97
1.0	0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.5	0.00	0.36	0.62	0.31	0.11	0.35	0.26	0.19
1.5	0.05	0.43	0.72	0.43	0.20	0.42	0.33	0.25
1.5	0.10	0.59	0.88	0.69	0.44	0.62	0.53	0.44
1.5	0.15	0.78	0.97	0.89	0.72	0.83	0.76	0.70
1.5	0.20	0.91	0.99	0.97	0.91	0.95	0.92	0.89
1.5	0.25	0.98	1.00	1.00	0.98	0.99	0.98	0.97
1.5	0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2.0	0.00	0.71	0.89	0.70	0.45	0.75	0.68	0.61
2.0	0.05	0.74	0.91	0.75	0.52	0.78	0.72	0.65
2.0	0.10	0.82	0.95	0.85	0.68	0.85	0.80	0.75
2.0	0.15	0.90	0.98	0.93	0.84	0.92	0.89	0.86
2.0	0.20	0.96	0.99	0.98	0.94	0.97	0.96	0.94
2.0	0.25	0.99	1.00	0.99	0.98	0.99	0.99	0.98
2.0	0.30	1.00	1.00	1.00	1.00	1.00	1.00	0.99
2.5	0.00	0.89	0.97	0.89	0.75	0.92	0.89	0.85
2.5	0.05	0.90	0.97	0.90	0.77	0.93	0.90	0.87
2.5	0.10	0.92	0.98	0.94	0.84	0.95	0.93	0.90
2.5	0.15	0.95	0.99	0.97	0.91	0.97	0.95	0.94
2.5	0.20	0.98	1.00	0.99	0.96	0.98	0.98	0.97
2.5	0.25	0.99	1.00	0.99	0.98	0.99	0.99	0.99
2.5	0.30	1.00	1.00	1.00	0.99	1.00	1.00	1.00
3.0	0.00	0.96	1.00	0.96	0.89	0.97	0.96	0.95
3.0	0.05	0.96	1.00	0.96	0.90	0.97	0.96	0.95
3.0	0.10	0.97	1.00	0.97	0.92	0.98	0.97	0.96
3.0	0.15	0.98	1.00	0.98	0.95	0.99	0.98	0.97
3.0	0.20	0.99	1.00	0.99	0.97	0.99	0.99	0.98
3.0	0.25	0.99	1.00	1.00	0.99	1.00	0.99	0.99
3.0	0.30	1.00	1.00	1.00	0.99	1.00	1.00	1.00

Table 3. Probabilities of Non-Proficient Rating Under Two Different A-Score Criteria (Compared to Current Four-Round Criterion)

ρ	$\|B\|$	$1-p_4$	$A_{1,1} = \|M_z\| + S_z$ C=3	C=3.5	C=4	$A_{2,2} = M_z^2 + S_z^2$ C=6	C=6.5	C=7
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0	0.10	0.00	0.01	0.00	0.00	0.00	0.00	0.00
1.0	0.15	0.01	0.19	0.01	0.00	0.03	0.01	0.00
1.0	0.20	0.15	0.76	0.25	0.02	0.44	0.31	0.20
1.0	0.25	0.65	0.99	0.81	0.31	0.94	0.89	0.81
1.0	0.30	0.97	1.00	0.99	0.85	1.00	1.00	0.99
1.5	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
1.5	0.05	0.00	0.04	0.00	0.00	0.01	0.01	0.00
1.5	0.10	0.02	0.25	0.05	0.00	0.07	0.04	0.02
1.5	0.15	0.13	0.65	0.28	0.06	0.33	0.24	0.16
1.5	0.20	0.42	0.93	0.69	0.32	0.74	0.65	0.55
1.5	0.25	0.79	0.99	0.94	0.72	0.96	0.93	0.90
1.5	0.30	0.97	1.00	1.00	0.95	1.00	1.00	0.99
2.0	0.00	0.07	0.18	0.04	0.01	0.25	0.18	0.12
2.0	0.05	0.09	0.31	0.10	0.02	0.30	0.22	0.16
2.0	0.10	0.18	0.61	0.32	0.11	0.47	0.38	0.30
2.0	0.15	0.36	0.86	0.63	0.34	0.71	0.63	0.55
2.0	0.20	0.63	0.97	0.87	0.65	0.90	0.86	0.80
2.0	0.25	0.85	1.00	0.97	0.88	0.98	0.97	0.95
2.0	0.30	0.97	1.00	1.00	0.98	1.00	1.00	0.99
2.5	0.00	0.33	0.59	0.29	0.10	0.71	0.64	0.56
2.5	0.05	0.36	0.67	0.39	0.17	0.74	0.67	0.60
2.5	0.10	0.46	0.82	0.60	0.36	0.82	0.77	0.70
2.5	0.15	0.61	0.94	0.81	0.61	0.91	0.87	0.83
2.5	0.20	0.78	0.98	0.94	0.82	0.97	0.95	0.93
2.5	0.25	0.90	1.00	0.99	0.94	0.99	0.99	0.98
2.5	0.30	0.97	1.00	1.00	0.99	1.00	1.00	1.00
3.0	0.00	0.63	0.86	0.64	0.38	0.93	0.90	0.86
3.0	0.05	0.66	0.89	0.69	0.45	0.93	0.91	0.87
3.0	0.10	0.71	0.93	0.80	0.61	0.95	0.93	0.91
3.0	0.15	0.79	0.97	0.91	0.78	0.98	0.96	0.95
3.0	0.20	0.88	0.99	0.97	0.91	0.99	0.98	0.98
3.0	0.25	0.94	1.00	0.99	0.97	1.00	0.99	0.99
3.0	0.30	0.98	1.00	1.00	0.99	1.00	1.00	1.00

Table 4. Comparison of A_1-Score Criteria with Unequal Weights (a,b) for Critical Value C = 3.5 and a + b = 2

ρ	$\lvert B \rvert$	$A_{1,1} = a\lvert M_z \rvert + bS_z$		
		a = 2/3 b = 4/3	a = 1 b = 1	a = 4/3 b = 2/3
1.0	0.00	0.00	0.00	0.00
1.0	0.05	0.00	0.00	0.00
1.0	0.10	0.00	0.00	0.00
1.0	0.15	0.00	0.01	0.05
1.0	0.20	0.05	0.25	0.57
1.0	0.25	0.32	0.81	0.97
1.0	0.30	0.75	0.99	1.00
1.5	0.00	0.01	0.00	0.00
1.5	0.05	0.02	0.00	0.00
1.5	0.10	0.09	0.05	0.04
1.5	0.15	0.28	0.28	0.30
1.5	0.20	0.57	0.69	0.75
1.5	0.25	0.84	0.94	0.97
1.5	0.30	0.96	1.00	1.00
2.0	0.00	0.24	0.04	0.01
2.0	0.05	0.32	0.10	0.04
2.0	0.10	0.52	0.32	0.19
2.0	0.15	0.74	0.63	0.52
2.0	0.20	0.89	0.87	0.83
2.0	0.25	0.97	0.97	0.97
2.0	0.30	0.99	1.00	1.00
2.5	0.00	0.70	0.29	0.07
2.5	0.05	0.74	0.39	0.15
2.5	0.10	0.84	0.60	0.37
2.5	0.15	0.92	0.81	0.66
2.5	0.20	0.97	0.94	0.87
2.5	0.25	0.99	0.99	0.97
2.5	0.30	1.00	1.00	1.00
3.0	0.00	0.92	0.64	0.23
3.0	0.05	0.93	0.69	0.31
3.0	0.10	0.96	0.80	0.52
3.0	0.15	0.98	0.91	0.74
3.0	0.20	0.99	0.97	0.89
3.0	0.25	1.00	0.99	0.97
3.0	0.30	1.00	1.00	0.99

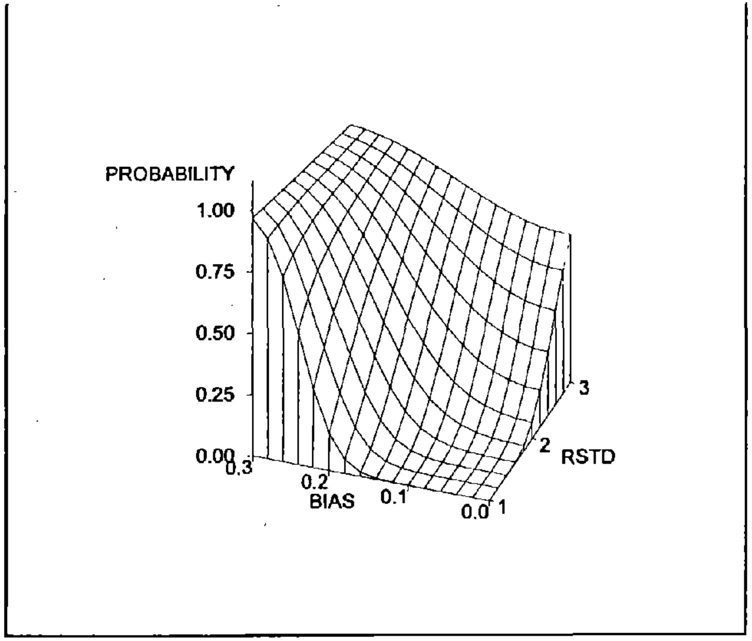

Figure 1. Current four-round criterion

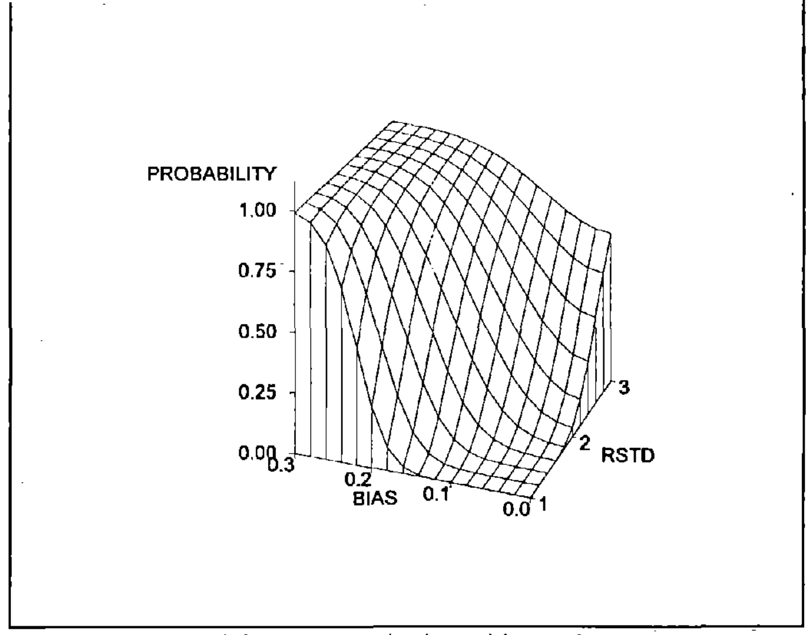

Figure 2. Proposed four-round criterion with equal weight: Probability = Prob($|M_z| + S_z > 3.5$)

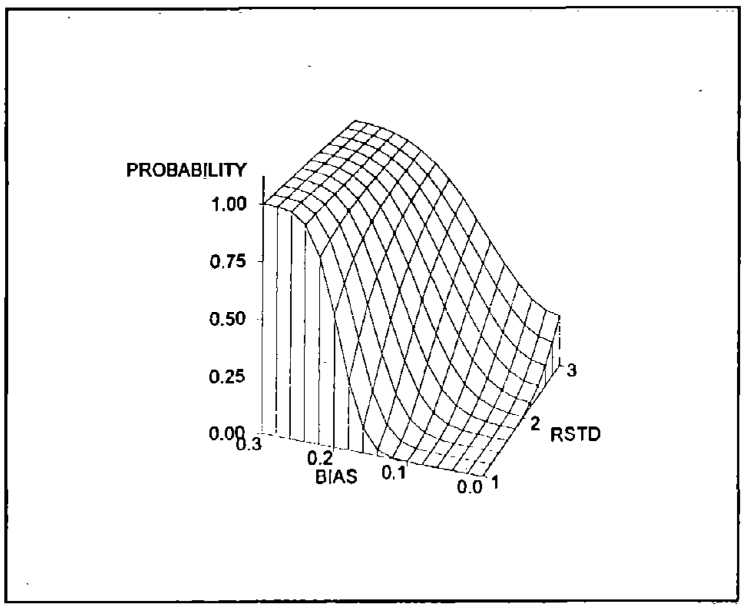

Figure 3. Proposed four-round criterion with more weight on bias:
Probability = Prob($(4|M_z| + 2S_z)/3 > 3.5$)

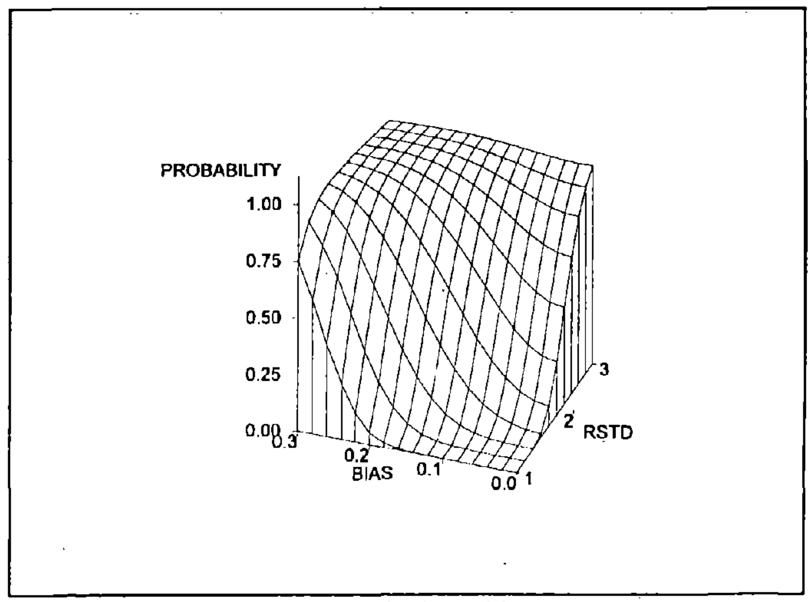

Figure 2. Proposed four-round criterion with more weight on precision:
Probability = Prob($(2|M_z| + 4S_z)/3 > 3.5$)

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO.	2.	PB97-147128

4. Title and Subtitle Laboratory Performance Criteria in the Environmental Lead Proficiency Analytical Testing (ELPAT) Program	5. Report Date 1996/12/00
	6.
7. Author(s) Schlecht, P. C., and R. Song	8. Performing Organization Rept. No.
9. Performing Organization Name and Address Division of Physical Sciences and Engineering, NIOSH, U.S. Department of Health and Human Services, Cincinnati, Ohio	10. Project/Task/Work Unit No.
	11. Contract (C) or Grant(G) No. (C) (G)
12. Sponsoring Organization Name and Address	13. Type of Report & Period Covered
	14.

15. Supplementary Notes

16. Abstract (Limit: 200 words) This report examines the rating criteria used in the Environmental Lead Proficiency Analytical Testing (ELPAT) Program. ELPAT employs simple criteria based on the number of acceptable results to rate a laboratory's performance. In this study, the statistical power of the current criteria was calculated and compared with alternative criteria based on z-scores. The results indicated that the alternative criteria are more powerful in detecting poor laboratory performance and more flexible in detecting bias or poor precision. The alternative criteria can be easily adjusted for various needs and the estimates of bias and precision of each participant can be obtained using the z-scores.

17. Document Analysis a. Descriptors

b. Identifiers/Open-Ended Terms NIOSH-Publication, NIOSH-Author, Chemical-analysis, Lead-compounds, Analytical-chemistry, Analytical-methods, Laboratory-testing, Statistical-analysis

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages 22
	22. Security Class (This Page)	22. Price

(See ANSI-Z39.18) *See Instructions on Reverse* **OPTIONAL FORM 272(4-77)** (Formerly NTIS-35) Department of Commerce