American Statistical Association

1999
Proceedings
of the
Section on Physical and Engineering Scienc

Papers presented at the

1999 Spring Research Conference on Statistics in Industry and Technology,
Minneapolis, Minnesota, June 2–4, 1999

Annual Meeting of the American Statistical Association,
Baltimore, Maryland, August 8–12, 1999,

1999 Fall Technical Conference,
Houston, Texas, October 14 & 15, 1999

all under the sponsorship of
the Section on Physical and Engineering Sciences

American Statistical Association ■ 1429 Duke Street, Alexandri

COMPARISON OF PERFORMANCE RATING CRITERIA IN PROFICIENCY TESTING PROGRAMS

Ruiguang Song[a] **and Paul C. Schlecht**[b]

[a]HGO Technology, Inc./NIOSH, [b]National Institute for Occupational Safety and Health

Ruiguang Song, DPSE/NIOSH, MS: R-7, 4676 Columbia Parkway, Cincinnati, OH 45226, USA

Key Words: Proficiency Testing, Rating Criteria, Bias, Precision, Outlier, Z-Score, Running Performance Index, Sample Size, Statistical Power.

Abstract

Two major proficiency testing programs in the industrial hygiene field are compared with regard to their power to detect laboratories having a large bias and poor precision. Although the two programs are similar in many aspects, such as the analytes tested, frequency of test rounds, and number of samples used in each test round, their evaluation procedures and rating criteria are quite different.

To expand the substances used in the proficiency testing and reduce the proficiency testing cost in each program, cooperation between the two programs will begin in 2000. Cooperation will involve exchange of samples, coordination of logistics, and harmonization of analytical methods, but will permit each program to continue its rating criterion. This study provides detailed power comparison of the two programs on rating laboratories at various bias and precision levels.

In the Workplace Analysis Scheme for Proficiency (WASP) program, participating laboratories are classified into three categories: "*average*", "*better than average*", and "*worse than average*". Laboratories in the Proficiency Analytical Testing (PAT) program are rated as "*proficient*" or "*non-proficient*". If a "*non-proficient*" rating in PAT is compared to a "*worse than average*" classification in WASP, the study shows that WASP is more sensitive in detecting laboratories with poor performance. That is, the chance for a laboratory having a large bias or low precision to be rated "*non-proficient*" in PAT is less than the chance of being classified as "*worse than average*" in WASP.

Although the PAT criterion is simpler, it is not as powerful because PAT converts each quantitative laboratory result to a qualitative value. The effect of information loss can be measured by the number of samples required to maintain the same power in both programs. The required number of samples for WASP is only 60%-80% the number of samples used in PAT.

To improve the PAT rating criterion a modified rating criterion based on z-scores has been previously proposed by Schlecht and Song (1997). This study demonstrates that the modified PAT criterion is equivalent to the WASP criterion.

1. Introduction

Proficiency testing programs are used to evaluate laboratory performance in a wide variety of fields. Two well-known proficiency testing programs in the industrial hygiene field are the United States-based Proficiency Analytical Testing (PAT) Program[1] and the United Kingdom-based Workplace Analysis Scheme for Proficiency (WASP) Program.[2]

The PAT program was started in 1972. It is used by the American Industrial Hygiene Association (AIHA) to accredit laboratories analyzing hazardous substances in workplace air, including various metals, organics, silica, and asbestos. Currently there are about 1,300 laboratories, mainly in the United States and Canada, participating in this program.

In 1988, the WASP Program was started in the United Kingdom. It is used by the United Kingdom Accreditation Service (UKAS) to accredit laboratories analyzing hazardous substances in the workplace and ambient air, including various metals, organics, silica, isocyanates and formaldehyde. About 300 laboratories, mainly from United Kingdom and other European countries, participate in this program.

Both programs send samples to laboratories on a quarterly basis. Four samples of each selected analyte are sent to each participating laboratory. Performance of each participating laboratory is evaluated based on results from the last four or five rounds. The PAT program rates laboratories into two categories: "*proficient*" or "*non-proficient*." While the WASP program classifies laboratories into three categories: 1 as "*better than average*", 2 "*average*", and 3 "*worse than average*."

To expand the substances used in the proficiency testing and reduce the proficiency testing cost in each program, cooperation between the two programs will begin in 2000.[3] Although the two programs are similar, their rating criteria are quite different. In this study, we compare the two programs, evaluate the performance rating criteria used in each program, and compute the probability that a laboratory is rated as "*non-proficient*" in PAT and the probability that the same laboratory is classified in category 3 in WASP.

To simplify the comparison, we only consider the simplest case, assuming that true concentrations are known and all results are independent. In the next section, we briefly describe the rating procedures in the two programs. Power associated with each program is presented in section 3. In section 4, we compare the two programs by the required sample size for each program to maintain minimum acceptable power. This gives a measure of the difference between the rating criteria in the two programs. In both the PAT and WASP programs, rating criteria are designed to permit occasional poor laboratory performance or possible sample contamination in a round by excluding data. The effects of excluding data on rating power are studied in section 5. In section 6, we show that the modified criterion proposed by Schlecht and Song[4] is statistically equivalent to the WASP criterion. Finally, summary results and some thoughts on the implications of this work on proficiency test programs are discussed in section 7.

2. Performance Rating Criteria in PAT & WASP

In both the PAT and WASP programs, four samples are distributed to each participating laboratory each quarter. Reported results from laboratories are then converted to z-scores in PAT or to standardized values in WASP. Let x_{ij} be the reported result from a laboratory for the j^{th} sample in the i^{th} round. Then the standardized value is given by $y_{ij} = (x_{ij} - \mu_{ij}) / \mu_{ij}$ and the z-score is defined as $z_{ij} = (x_{ij} - \mu_{ij}) / \sigma_{ij}$, where μ_{ij} and σ_{ij} are the reference mean and the reference standard deviation.

In PAT, each result is rated as "*acceptable*" or "*not acceptable*" (called an outlier). A laboratory is rated as "*proficient*" or "*non-proficient*" based on the total number of outliers in the last four rounds. A result is an outlier if the absolute value of its corresponding z-score is greater than 3. A laboratory is rated as "*proficient*" if it has either (A) no more than 4 outliers (out of 16 results) in the last four rounds, or (B) no outlier in the last two rounds.

In WASP, a laboratory's performance is evaluated based on the running performance index (*RPI*) that is defined as the average of the four best performance indexes in the last five rounds:

$$RPI = \left(\sum_{i=1}^{5} PI_i - \max\{PI_i\}\right) / 4 ,$$

where PI_i is the performance index for the *i*-th round: $PI_i = \sum_{j=1}^{4} y_{ij}^2 / 4$. If a laboratory's *RPI* is in the interval $(0.432RSD_0^2, 1.8RSD_0^2)$, then the laboratory is categorized as "*average*", where $RSD_0 = \sigma_0/\mu_0$ is a pre-specified relative standard deviation, which is different for different analytes. If $RPI < 0.432RSD_0^2$, the laboratory is identified as "***better than average***." If $RPI > 1.8RSD_0^2$, the laboratory is classified as "*worse than average*."

We have seen that the two programs have very different rating criteria. While our ultimate goal is to adopt identical rating criteria, this may not be immediately possible for several reasons. First, the rating scheme may be mandated by regulation. Second, the rating criteria may be the basis of mutual recognition agreements among accrediting organizations and governmental agencies. Third, criteria may be used differently in different programs. In some, a "*non-proficient*" rating results in immediate loss of accreditation and loss of business by the laboratory, whereas other programs may permit investigations of laboratory operations before this occurs. Therefore, in some instances a type I error must be very small, whereas in other instances a moderate type I error can be tolerated. Fourth, many proficiency test programs must address issues that are unique to a particular type of analysis and to the type of participating laboratory, which makes adoption of some rating criteria not feasible. For example, AIHA operates a diffusive monitor proficiency test program for various organics where the cost of samples makes it not feasible to operate more than two rounds (8 samples) per year. Similarly, another AIHA program focuses on proficiency testing of microscopists performing fiber counting of asbestos and other fibers. These microscopists are mobile and must deal with the unique problems this presents for logistics. Fifth, since there are many proficiency test programs, adoption of multiple rating criteria to satisfy multiple cooperating proficiency test program may be confusing for laboratories. Therefore, until a single proficiency test protocol can be universally agreed upon by the many proficiency test programs, a comparison of criteria, based on power and simple modifications to criteria, to made the criteria similar, is useful. It is the first step towards broader cooperation and harmonization among programs.

3. Power Comparison

To compare the rating criteria in the two programs, we selected a laboratory and assumed that this laboratory had a consistent performance with a constant bias B (relative difference from the true value) and a constant precision *TRSD* (defined as the standard deviation divided by the true value). Then, we compared the chance for this laboratory to be rated as "*non-proficient*" according to the PAT criterion with its chance to be classified in category 3 under the WASP criterion. Since the WASP program uses the best four performance rounds (out of five) in its criterion and the PAT program has an additional two-round criterion, there are explicit formulas to calculate the rating power associated with each criterion. Therefore, a simulation study was conducted for various bias and precision levels. Simulation results from 10,000 replicates are shown in Table 1 for PAT and Table 2 for WASP.

Table 1. Probabilities for a laboratory to be rated non-proficient using the **PAT** criterion.
(Simulation results from 10000 replicates, the reference $TRSD_0$=0.06)

	TRSD										
BIAS	***0.05***	***0.06***	***0.07***	***0.08***	***0.09***	***0.10***	***0.11***	***0.12***	***0.13***	***0.14***	***0.15***
0.00	0.000	0.000	0.000	0.000	0.001	0.005	0.018	0.051	0.108	0.195	0.296
0.01	0.000	0.000	0.000	0.000	0.001	0.005	0.019	0.054	0.113	0.196	0.299
0.02	0.000	0.000	0.000	0.000	0.001	0.006	0.022	0.062	0.123	0.207	0.306
0.03	0.000	0.000	0.000	0.000	0.002	0.008	0.031	0.070	0.138	0.222	0.323
0.04	0.000	0.000	0.000	0.001	0.003	0.014	0.042	0.089	0.158	0.247	0.347
0.05	0.000	0.000	0.000	0.002	0.006	0.023	0.057	0.113	0.191	0.285	0.383
0.06	0.000	0.000	0.000	0.003	0.015	0.040	0.083	0.147	0.234	0.330	0.424
0.07	0.000	0.000	0.002	0.009	0.030	0.068	0.120	0.197	0.288	0.381	0.470
0.08	0.000	0.001	0.005	0.023	0.056	0.107	0.175	0.261	0.350	0.433	0.521
0.09	0.000	0.003	0.018	0.048	0.100	0.166	0.247	0.333	0.414	0.499	0.579
0.10	0.002	0.013	0.044	0.097	0.167	0.246	0.329	0.412	0.491	0.566	0.633
0.11	0.007	0.039	0.097	0.178	0.264	0.345	0.427	0.504	0.569	0.635	0.695
0.12	0.031	0.097	0.191	0.288	0.377	0.460	0.529	0.595	0.655	0.707	0.753
0.13	0.097	0.212	0.323	0.427	0.505	0.576	0.634	0.687	0.731	0.770	0.805
0.14	0.238	0.376	0.487	0.573	0.639	0.691	0.732	0.768	0.797	0.828	0.852
0.15	0.454	0.573	0.658	0.718	0.758	0.787	0.814	0.837	0.856	0.877	0.893

Note: Probabilities less than 0.05 are shaded.

Table 2. Probabilities for a laboratory to be classified in category 3 based on the **WASP** criterion.
(Simulation results from 10000 replicates, the reference $TRSD_0$=0.06)

	TRSD										
BIAS	***0.05***	***0.06***	***0.07***	***0.08***	***0.09***	***0.10***	***0.11***	***0.12***	***0.13***	***0.14***	***0.15***
0.00	0.000	0.002	0.037	0.181	0.413	0.646	0.812	0.907	0.956	0.980	0.990
0.01	0.000	0.003	0.044	0.193	0.424	0.659	0.820	0.910	0.956	0.980	0.991
0.02	0.000	0.007	0.065	0.236	0.465	0.688	0.833	0.920	0.959	0.982	0.992
0.03	0.001	0.021	0.109	0.304	0.533	0.734	0.858	0.929	0.965	0.983	0.992
0.04	0.007	0.056	0.197	0.410	0.623	0.789	0.887	0.942	0.972	0.987	0.994
0.05	0.038	0.149	0.337	0.540	0.721	0.842	0.916	0.958	0.979	0.989	0.994
0.06	0.150	0.321	0.513	0.681	0.812	0.894	0.942	0.971	0.985	0.991	0.995
0.07	0.397	0.557	0.702	0.811	0.884	0.933	0.964	0.981	0.989	0.994	0.997
0.08	0.705	0.784	0.850	0.902	0.939	0.966	0.980	0.990	0.994	0.997	0.998
0.09	0.909	0.924	0.943	0.959	0.973	0.984	0.991	0.995	0.997	0.998	0.999
0.10	0.985	0.983	0.984	0.987	0.990	0.994	0.995	0.998	0.998	0.999	0.999
0.11	0.998	0.997	0.996	0.996	0.996	0.997	0.999	0.999	0.999	0.999	1.000
0.12	1.000	1.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: Probabilities less than 0.05 are shaded.

From the two tables, we see that a laboratory is much less likely to get a non-proficient rating from PAT than a category III worse than average rating from WASP. This shows that the WASP criterion is much stricter than the PAT criterion and is the main reason why the PAT program has a lower percentage[5](<10%) of laboratories rated as non-proficient than that (>20%) classified as worse than average in the WASP program[6].

4. Sample Size Comparison

A stricter criterion may or may not be a more powerful criterion. To study the power of a rating criterion, we view a performance rating as a statistical hypothesis testing. The hypotheses specify a laboratory's performance characteristics that are determined by the combination of bias and precision. The null hypothesis specifies the characteristics for a laboratory that deserves a good rating. The expected precision for a good laboratory depends on the analyte under test and has been explicitly specified for each analyte in WASP. In PAT, it is estimated from reference laboratories' results. The alternative hypothesis characterizes laboratories with unacceptable performance.

To compare the power of two rating criteria, it is necessary to specify the null hypothesis, set the two tests at the same level of type I error, specify the alternative hypothesis, and compare the type II error evaluated at the alternative hypothesis. Since the power is associated with sample size, we may compare the required sample sizes that make the two tests have the same type II error.

To simplify the comparison, we only consider the criterion using four-round results in each program so we can calculate the power without simulation. Also, we assume that all results are independent with an identical normal distribution. Under these assumptions, we have $z_{ij} \sim N(B/TRSD_0, TRSD^2/TRSD_0^2)$ and $y_{ij} \sim N(B, TRSD^2)$, where $TRSD_0$ is the expected value of $TRSD$. In this case, the WASP rating score RPI has a χ^2 distribution with a degrees of freedom N (=16, the total number of samples tested). The PAT rating score S is the total number of outliers. It has a binomial distribution $Bin(N, p)$, where $p = \Pr(|z_{ij}| > k)$ is the probability of getting an outlier and $k=3$ in PAT.

Let L be the number of outliers permitted for getting a proficient rating. Then the chance for a laboratory to be rated as non-proficient is $P_{NP} = \Pr(S > L)$. The two parameters L and k can be adjusted so that $P_{NP} = \alpha$ under the null hypothesis H_0: $B=0$ and $TRSD=TRSD_0$.

In WASP, we can adjust the limit for the running performance index, denoted by H, such that the chance for a laboratory to be categorized as "*worse than average*" is also α: $\Pr(RPI > H) = \alpha$ under the null hypothesis. Note that $RPI \sim TRSD^2 \times \chi^2(N,D)$, where $D = N \times (B/TRSD)^2$ is the non-central parameter. The critical value H can be determined by $H = TRSD_0^2 \times \chi^2(\alpha, N)$. Using this H, $\Pr(RPI > H)$ can be calculated.

We set $\alpha=0.025$, $TRSD_0=0.06$, and the alternative hypothesis H_1: $B=0.05$ and $TRSD = 2 \times TRSD_0$. Given the sample size N in PAT, we calculate the sample size N' required in WASP to maintain the same power as in PAT at the alternative hypothesis. Calculation results are shown in Table 3, where $\lambda = L/N$ is the outlier rate allowed for a laboratory to be rated as proficient in PAT, k is selected to give the PAT a type I error $\alpha=0.025$ to match the type I error in WASP, β_{PAT} is the type II error associated with the PAT criterion, $\beta_{WASP(N)}$ and $\beta_{WASP(N')}$ are the type II errors associated with the WASP criterion for sample sizes N and N', respectively. From this table, we can see that there are about 25%~40% information lost by converting a continuous variable to a binary variable in PAT. The loss depends on the outlier rate λ and the total number of samples N.

5. Poor Laboratory Performance or Suspect Samples on a Given Round

Both programs have a mechanism to permit laboratories that are rated "*non-proficient*" or "*worse than average*" due to poor performance or receipt of suspect samples on a given round to improve ratings. In PAT, this is done by using the two-round criterion. If the laboratory has no outliers on the last two rounds, the laboratory is rated proficient even though previous rounds have many outliers. In WASP, one extra round's results are used and the worst round results are eliminated from the rating. To see the effects of these additional criteria, we show the simulation results in Table 4 for the case bias B=0. Column P4 gives the probabilities for a laboratory getting a "*non-proficient*" rating under the PAT four-round criterion. Probabilities in column +P2 show the decreased chance for a laboratory getting a "*non-proficient*" rating, in other words, the extra chance to get a proficient rating. Similarly, the column W4 gives the probabilities for a laboratory to get a "*worse than average*" rating under the WASP criterion using four-round results only. Probabilities in column +W5 is the decreased chance for a laboratory getting a worse than average rating.

Table 3. Required sample sizes to maintain the power at H_1.

λ	N	m	k	p	β_{PAT}	$\beta_{WASP(N)}$	N'	$\beta_{WASP(N')}$	N'/N
1/4	12	3	1.65	0.449	0.136	0.040	8	0.117	0.667
	16	4	1.60	0.463	0.070	0.013	10	0.069	0.625
	20	5	1.56	0.474	0.035	0.004	13	0.030	0.650
	24	6	1.53	0.483	0.017	0.001	15	0.017	0.625
	28	7	1.51	0.490	0.008	0.000	18	0.007	0.643
	32	8	1.49	0.495	0.004	0.000	21	0.003	0.656
1/5	10	2	1.83	0.400	0.168	0.069	7	0.153	0.700
	15	3	1.76	0.418	0.070	0.017	10	0.069	0.667
	20	4	1.71	0.431	0.028	0.004	14	0.023	0.700
	25	5	1.68	0.441	0.011	0.001	17	0.010	0.680
	30	6	1.65	0.449	0.004	0.000	21	0.003	0.700
1/6	12	2	1.92	0.378	0.110	0.040	9	0.090	0.750
	18	3	1.85	0.395	0.036	0.007	13	0.030	0.722
	24	4	1.80	0.408	0.011	0.001	17	0.010	0.708
	30	5	1.77	0.417	0.003	0.000	21	0.003	0.700

Table 4. Probability of a laboratory receiving a "*proficient*" rating from the PAT program or a "*worse than average*" rating from the WASP program (Simulation results of 10000 replicates, bias B=0)

	PAT			WASP		
TRSD	P4	+P2	Overall	W4	+W5	Overall
0.05	0.0000	0.0000	0.0000	0.0004	0.0003	0.0001
0.06	0.0000	0.0000	0.0000	0.0239	0.0220	0.0019
0.07	0.0000	0.0000	0.0000	0.1725	0.1353	0.0372
0.08	0.0001	0.0000	0.0001	0.4346	0.2536	0.1810
0.09	0.0006	0.0000	0.0006	0.6840	0.2710	0.4130
0.10	0.0046	0.0000	0.0046	0.8491	0.2032	0.6459
0.12	0.0514	0.0001	0.0513	0.9720	0.0654	0.9066
0.15	0.2979	0.0017	0.2962	0.9978	0.0075	0.9903
0.20	0.7553	0.0029	0.7524	1.0000	0.0001	0.9999

For the PAT program, the two-round criterion does not increase the laboratory's chance to be rated proficient if this laboratory has not improved its performance in the last two rounds. Only those labs whose performance was poor and then improved in the last two rounds can benefit from the two round criterion.

In WASP, the chance based on 4 round criterion differs from the chance based on the best 4 out of the last 5 rounds. The chance could be 27% higher without performance improvement. This benefits laboratories that have performance with moderate bias and precision.

6. A Modified PAT Criterion

Schlecht and Song[4] proposed a criterion in 1997 to improve the power of the existing criterion used in the Environment Lead Proficiency Analytical Testing (ELPAT) program and the PAT program (both share the evaluation procedure and rating criterion). The proposed criterion is based on the combination of mean and standard deviation of z-scores: $A = a|\bar{z}|^r + b\sigma_z^s$, where $\bar{z}$ and σ_z are the mean and standard deviation of z-scores.

Notice that $y_{ij} = TRSD_0 \times z_{ij}$ and

$$RPI_4 = \sum_{i=1}^4 PI_i/4 = \sum_{i=1}^4 \sum_{j=1}^4 y_{ij}^2/16$$
$$= TRSD_0^2 \times \sum_{i=1}^4 \sum_{j=1}^4 z_{ij}^2/16 = TRSD_0^2 \times \left(\bar{z}^2 + \hat{\sigma}_z^2\right)$$

If we set $r = s = 2$ and $a = b = TRSD_0^2$, then the modified rating score A is exactly the same as the running performance index over four results. So the modified criterion must have the same power as the WASP criterion in detecting laboratories with poor performance.

The modified criterion is flexible in determining the power to detect bias or poor precision. To be sensitive in detecting bias, we can set a large value for *a* and a relatively small value for *b*. Similarly, by setting a large value for *b* and a relatively small value for *a*, we may have a criterion more sensitive to detect laboratories with low precision.

The modified criterion is also meaningful. The average z-score is a measure of bias and the standard deviation of the z-score is a measure of precision. The rating score A is simply a combination of bias and precision estimates.

7. Summary and Discussions

This study compares the power of rating criteria used in the two largest industrial hygiene laboratory proficiency test programs, the US-based PAT Program and the UK-based WASP Program. These programs will begin an exchange of samples in 2000, but have opted to continue to use their current rating criteria. The WASP criterion is more powerful than the PAT criterion. This is because there is an information loss in PAT when it transfers a continuous variable to a binary variable. To measure the loss, the required sample sizes for maintaining the same power are compared. The WASP criterion is more complex, but requires 20% to 40% fewer samples to achieve the same power as PAT. For best use of PAT data, no results should be grouped or truncated.

The two-round criterion in PAT encourages laboratories to quickly improve their performance, without changing the ratings of other participating laboratories. On the other hand, the WASP best 4 out of 5 criterion significantly changes the rating of participating laboratories, thereby, benefiting laboratories with modest bias and imprecision.

To improve the efficiency of using data in PAT, we recommend using z-scores in the performance rating, instead of the number of outliers. The modified criterion proposed by Schlecht and Song[4] can be easily adopted. It is not only more powerful than the current PAT criterion and comparable to the WASP criterion, it is also meaningful and flexible in adjusting its sensitivity to a laboratory's bias or precision. Such a modification to minimize the loss of information is important in situations where a small number of samples are used to rate laboratory performance, such as the PAT proficiency testing of diffusive monitors where only two rounds (8 samples) are used to rate performance. The change would be easy for laboratories to understand because it is based on z-scores that are already reported to PAT participating laboratories.

In this study, we assume that true concentrations are known. Since in both programs, concentrations are unknown, and estimated by the average of all laboratories' results. This can make some difference in the power comparison. However, the differences should be small because the number of participating laboratories is large (50-1000).

This study compared the power of two large proficiency test programs and suggested simple changes to criteria to harmonize them. Comparing proficiency test rating criteria based on power tables of probabilities versus laboratory bias and precision or plots of this same information, sometimes referred to as operating characteristic (OC) curves, could be a means to compare multiple proficiency test programs that have unique criteria. Such could be an important first step towards expanded cooperation among programs and an interim solution to harmonizing criteria among multiple proficiency test programs. An interim solution to harmonizing criteria is useful because it is difficult to adopt new criteria when criteria are used or mandated by various laboratory accreditation organizations and government agencies, and participating laboratories are familiar with existing criteria.

References

[1] 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 94 - 102).*

[2] H. M. Jackson and N. G. West (1993). Initial Experience with the Workplace Analysis Scheme for Proficiency (WASP). *Ann. Occup. Hyg., 36(5), 545-561.*

[3] Martin Harper (1999). Laboratory Accreditation: PATs Across the Pond. *The Synergist, 10(8), 13-14.*

[4] Paul Schlecht and Ruiguang Song (1997). Laboratory performance criteria in the Environmental Lead Proficiency Analytical Testing (ELPAT) program. *DHHS(NIOSH) Pub. No. PB97-147128, Cincinnati, OH: National Institute for Occupational Safety and Health.*

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

[6] Barry Tylee (1995), The WASP News, Autumn issue, United Kingdom Health & Safety Executive, Shelfield.