# Examining the Whys and Hows of Proficiency Testing

## INTRODUCTION TO PROFICIENCY TESTING

The new international standard, ISO/IEC 17025 *General Requirements for the Competence of Testing and Calibration Laboratories*[1], references proficiency testing in the context of monitoring activities associated with testing and calibration. Per section 5.9, monitoring activities are to be conducted in such a manner as to provide data so that trends can be detected, and where practical, statistical techniques applied in the examining of this data. In addition, monitoring activities shall be planned and reviewed. But what is proficiency testing and what are its relevant properties?

- **Why perform monitoring activities if test/measurement equipment is routinely calibrated and maintained?**
- **What guidelines are used to govern monitoring activities?**
- **On what, specifically, are these monitoring activities being performed?**
- **What performance statistics are employed during data monitoring?**
- **What processes and activities are involved in the planning and review of monitoring activities?**

These are a few of the questions that will be addressed in this paper within the perspective of proficiency testing.

Proficiency testing is normally differentiated from interlaboratory comparisons by the independence of the laboratory assigning values to test artifacts. NCSL RP-15 *Guide for Interlaboratory Comparisons*[2] further relates that unless specifically stated, an **interlaboratory comparison** "does not necessarily demonstrate proficiency... and it is usually not accepted as meeting the proficiency requirements for accreditation." **Proficiency testing,** as literally defined in ISO/IEC Guide 43-1 *Proficiency testing by Interlaboratory Comparisons*[3], is a means used in the determination of laboratory testing and measurement performance. Stated a little differently, proficiency testing is a monitoring activity with the purpose of assessing the quality and uniformity of tests and measurements performed by a laboratory. The methodology for performance determination is established via the evaluation of measurements made on test artifacts or artifact standards. NCSL RP-15 defines an artifact standard as "a standard whose value or values are based on human skill and workmanship."

Test artifacts used for proficiency testing are normally devices (standards) exhibiting stable characteristics that are similar to units tested or calibrated by a laboratory. For many proficiency test applications, test material commonly referred to as reference material or standard reference material (SRM) or certified reference material (CRM) is the norm. This test material or substance is sufficiently stable and homogeneous regarding one or more of its properties, such as to enable it to be used in the assessment of measurement processes. In the context of this paper, test artifacts also refer to test materials as applicable.

Proficiency test artifact salient characteristics are assigned by a reputable, independent laboratory, i.e. a laboratory normally not participating in the proficiency test. Section 5.9 of ISO 17025, which references proficiency testing programs, is labeled *Assuring the Quality of Test and Calibration Results*. In looking up the meaning of "assure," the words promise, guarantee, pledge, declare, give surety, and comfort, are some definitions provided. Given these definitions, proficiency testing is basically an assessment activity used to demonstrate and/or determine a level of work competence as derived from measuring a test artifact and evaluating the results. The demonstrated work competence level is assumed to be representative of the germane work performed by a laboratory. As such, it is deemed a predictor of sorts in regards to the work a laboratory may be expected to perform. Consequently, the result of a laboratory satisfactorily completing a temperature proficiency test provides an assumption that the laboratory will continue to make satisfactory temperature measurements within the scope of the test. Proficiency test activities are performed independent of equipment calibration activities in order that proficiency test results may be used to determine the control or lack of control of measurement processes that employ and calibrate equipment.

It must be emphasized that a proficiency test is only applicable for the scope of the test being performed; assumptions made about other laboratory work should be avoided. The assumption that comparable results can be obtained as in a previous proficiency test using the same or similar test equipment, personnel, etc., but extending only marginally beyond the previous tests scope often proves invalid. The key to understanding proficiency testing is the term **demonstrated competence.** Demonstrated competence, as related to proficiency testing, is the documented evidence that a proficiency test was performed satisfactorily. The treatment and evaluation of proficiency test data is the means by which a laboratory's proficiency test competence is determined.

Standardization of the treatment and evaluation of proficiency test data has evolved over time and been a key factor in promoting proficiency test result acceptance by the international test and measurement community. ISO/IEC Guide 43-1 discusses methods for the treatment and evaluation of proficiency test data and is commonly listed by proficiency test providers as being the reference for the methodologies employed in modern test schemes. The International Laboratory Accreditation Co-operation (ILAC) publication, *Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes*[4] (from now on referred to as the "ILAC Guide"), defines the term provider as "a body (organization or firm, public or private) that undertakes the design and conduct of a proficiency test scheme."

## TEST ARTIFACT AND REFERENCE LABORATORY CONSIDERATIONS

Test artifacts and the measurements performed on them are the heart of proficiency testing. Financial restraints, resource limitations, and time considerations, are the main inhibitors in making onsite "over-the-shoulder" proficiency tests commonplace. Because of this, deployed test artifacts are usually the vehicle used to evaluate a laboratory's performance for a specific scope. The test artifact, in most circumstances, determines the level of uncertainty associated with a proficiency test. An example would be in the selection of a thermocouple versus a platinum resistance thermometer (PRT) for use in temperature proficiency testing. PRTs, by their composition and design, exhibit properties that allow lower uncertainties than those of thermocouples. Knowing which test artifacts are used in proficiency tests is a relatively simple means by which laboratory personnel can determine if a proficiency test is compatible with its own laboratory's capabilities and interests.

Proficiency test providers must be able to demonstrate that selected test artifacts are sufficiently stable and, if applicable, homogeneous for a particular proficiency test. This is particularly true given the reality that the handling and transportation of test artifacts between test participants is many times uncontrolled, even with the best of safeguards incorporated. Using proper packaging materials and shipping containers helps to ensure safe transit of artifacts and is especially critical for fragile items such as glass-sheathed PRTs that can easily be damaged or destroyed from rough handling. Special considerations for those units requiring battery power must also be made when shipping times might exceed battery life expectancy (many a horror story has been told of the reference voltage cells left stranded on a shipping dock only to have their battery backup expire and void an entire test program). Environmental factors must also be addressed for units susceptible to value shifts resulting from exposure to temperature, humidity, and/or pressure extremes. Proficiency test participants are normally required to inspect test artifacts upon receipt in order to identify any shipping damage. Proficiency test providers may require test participants to notify them when a test artifact has been shipped and/or received in order to avoid program delays and to flag missing artifacts at the earliest possible opportunity. Test participants may also be required to perform some type of assurance activity upon receipt of test artifacts so as to ensure their proper operation, i.e. measuring the resistive value of a thermistor used to monitor the ovenized enclosure of a Zener reference cell.

The selection of an independent laboratory for assignment of a proficiency test artifact's salient characteristics is key in establishing that test's evaluation criteria. The independent laboratory, commonly referred to as the **pivot laboratory (or reference laboratory** as referred to in this paper), is defined in NCSL RP-15 as "a laboratory that serves as the hub facility for an interlaboratory comparison program." A reference laboratory is selected based on several considerations. The major consideration for selection is a proven track record, established by means of a long-term demonstrated competence for making reliable measurements on devices similar to or the same as the test artifact that will be used in the proficiency test. A reference laboratory's proven track record can be surmised from several factors such as the laboratory's accreditation status, past proficiency tests, laboratory intercomparison results, national and/or international status, skills and training of laboratory personnel, the type of laboratory equipment and standards used, etc. Another consideration for selection of a reference laboratory is verification that its uncertainties are comparable to the uncertainties normally associated with the test artifact and the testing scheme. Lesser considerations include location, workload and turnaround times, cost, confidentiality, and conflict of interest issues.

Ethical considerations for selection of a reference laboratory, as well as choosing a proficiency test provider, are often ignored. Ethical considerations may be found in NCSL's RP-15 Section 4, **Basic Organization and Management.** Subsection 4.4 on ethics states that under no circumstances should the positions of coordinator, data processor, or pivot laboratory be filled by persons or organizations that could possibly profit from the knowledge of how a laboratory performs. Subsection 4.4.1 on participation states that designs that give one or more laboratories unfair advantage or intentionally exclude laboratories so as to gain advantages should not be used. Subsection 4.4.2 addresses anonymity and confidentiality, and is very clear regarding breaches of confidentiality, accidental or intentional, which may require termination of activities. These recommendations, though written in the context of interlaboratory comparisons, are just as applicable to proficiency testing, especially in cases where accreditation considerations are involved.

The ILAC Guide's section, **Confidentiality,** requires that proficiency test providers shall, unless waived by participants, keep the identity of participants to the minimum needed to conduct the test and that all information supplied by participants should be treated as confidential. The ILAC Guide also contains the section **Collusion and Falsification of Results,** which requires that proficiency test providers shall, when practical, design proficiency test schemes so that the opportunity for collusion and falsification of test results is minimized. Ethical considerations such as those contained in national and international guidelines and practices are necessary to minimize the possibility of competing interests obtaining unfair advantages by disclosing proficiency test results so as to tarnish a test participant's reputation

## PROFICIENCY TEST AND TEST SCHEME DEVELOPMENT

Proficiency test providers are tasked with designing and developing test schemes that are useful, practical, and desirable, as viewed from the perspective of a test participant. Test participants have the expectation that proficiency test providers will design and develop test schemes that comply with national and international guidelines and standards, as well as provide these tests in a manner that is timely, unambiguous, and error free. Typically, proficiency test providers survey prospective participants in order to determine the most prolific need for proficiency tests. With proficiency tests identified, target levels for uncertainties associated with each test must be established. These two activities have the greatest influence on the kind of test artifacts that may be used in a proficiency test.

After identifying the test artifacts to be used for a proficiency test, test scheme design and development can begin. The ILAC Guide defines proficiency test schemes as "interlaboratory comparisons designed and operated to assure laboratory performance in specified areas of testing, measurement or calibration." In addition, they note that "a scheme might cover a particular type of test or a number of tests on particular products, items and materials."

Some of the items proficiency test providers need to address when designing and developing a test scheme are:

- Nature and purpose of the test
- Participant and support information (logistics)
- Delivery schedules and shipping information
- Test artifact information (preparation, handling, uncertainty, etc.)
- Procedural and data acquisition information
- Analysis and reporting requirements

Proficiency test providers give general instructions and basic guidelines for conducting a proficiency test. In most cases, the selection of the actual test methods and the procedures for obtaining measurement data are left up to each participant laboratory. This is done to assure test results reflect not only the environment, equipment and personnel performing the measurements, but the actual methodologies the laboratory uses in its day-to-day operations. Test scheme design and development activities are conducted under the auspices of technical advisors often referred to as collaborators in the sense that many of them subcontract their expertise to proficiency test providers on an as-needed basis. The ILAC Guide defines a collaborator as "a body (organization or firm, public or private) that undertakes subcontracted activities for a proficiency testing scheme provider." Proficiency test providers recognize that they cannot be experts in all test and measurement fields and must rely on collaborative efforts to effectively design, develop and review proficiency test schemes. Technical advisors are selected based on their knowledge and skills relevant to the scope of the proposed proficiency test. Proficiency test providers are responsible for ensuring technical advisors and other key personnel have substantial documentation, i.e. training records, which attest to their knowledge and skills as needed to successfully perform their assigned tasks.

Technical advisors help determine:

- Test instructions and, if applicable, test methodologies
- Procedural and data acquisition requirements
- Analysis and reporting requirements
- Commentaries as applicable

Proficiency test providers rely on technical advisors to help ensure the technical validity of test results. Technical advisors are consulted on an as-needed basis throughout the life of a proficiency test in order to answer participant questions, monitor and review test results, and address any unforeseen problems that may arise. Frequently, as a result of the experience gained from conducting proficiency tests, technical advisors will change, add, or delete components of a test to increase efficiency, eliminate ambiguities, provide additional safeguards and guidance, and generally improve the overall testing scheme. Review activities, conducted by technical advisors, help to ensure that proficiency test schemes are dynamic in the sense that they are not only evaluated in terms of past test experience, but also are relevant to new technologies, test techniques, and analyses. It can be surmised that technical advisors are the cornerstones upon which proficiency test providers lay the bricks of their technical foundation and test services.

## BRIEF NOTE ON THE IMPORTANCE OF QUALITY SYSTEMS

Proficiency test providers, like all quality conscious service providers, should have a quality system in place as documented in their quality manual. Provisions within a proficiency provider's quality system should, as a starting point, address the topics contained within the ILAC Guide. This document, based upon the requirements of ISO Guide 43-1 and ISO/IEC 17025, is at the heart of the American Association for Laboratory Accreditation (A2LA) program for accrediting proficiency test providers. The ILAC Guide specifically states, as with ISO Guide 43-1 that "these ILAC requirements apply only to the use of interlaboratory comparisons for the purpose of proficiency testing, i.e. to determine the performance of individual laboratories for specific tests or measurements and to monitor laboratories continuing performance."

The ILAC Guide is divided into three major sections: **General, Management Systems Requirements, and Technical Requirement.** These major sections contain various topics, some common to management systems in general, such as **Corrective Action, Records, and Management Reviews,** while other sections are specific to proficiency testing such as **Organization and Scheme Design Logistics and Conduct of Proficiency Test Scheme.**

The ILAC Guide is particularly informative in regard to its Annex 2, *Cross-references to ISO 9000, ISO/IEC Guide 43 and ISO/IEC 17025.* Another useful publication focusing on proficiency testing and quality system requirements is the American Society for Testing and Materials (ASTM) E 1301-95, *Standard Guide for Proficiency Testing by Interlaboratory Comparisons*[5]. Section A2.1 of this publication states that "a documented quality assurance system should be established and maintained. It should outline the policies and procedures that exist to ensure the quality of interlaboratory testing services provided." This section clearly addresses the quality system procedures needed to adequately support proficiency test activities. Lower level documents provide the details and instructions needed to satisfactorily perform tasks using standardized procedures. As with any quality system, quality documents should be controlled and revised to ensure their continued viability.

## STATISTICAL DATA ANALYSIS: THE BASICS

As previously mentioned, the math used in the treatment and evaluation of proficiency tests has been standardized over time. The ILAC Guide covers the treatment and evaluation of proficiency test data in its Annex 1, *Statistical Methods for Treatment of Proficiency Test Data*. This annex specifies the following: "To assist providers of proficiency testing schemes, guidance on the selection and use of statistical procedures for the treatment of proficiency test data is given in Annex A of ISO/IEC Guide 43-1."

The major topics addressed by ISO/IEC Guide 43-1 Annex A, from now on referred to as "Annex A," are:

- Determination of the assigned values and its uncertainty
- Calculation of performed statistics
- Evaluation of performance
- Preliminary determination of test item homogeneity

Annex A states that it "gives general criteria for statistical techniques that can be applied as needed to guide applications" and "does not consider statistical techniques for analytical studies other than the treatment of proficiency test data." As such, Annex A references five of the most common procedures used when determining an assigned value for a proficiency test artifact.

The most common procedures used to assign test artifact values are (in order of relatively increasing uncertainty):

1) Known values (via an artifact's make up/formulation)
2) Certified reference values (derived via definitive methods)
3) Reference values (derived via comparison to traceable standard)
4) Consensus values from expert laboratories
5) Consensus values from participant laboratories

Assigned test artifact values, as stated in Annex A, should be determined so that participants can be evaluated fairly by selecting common comparison groups and using common assigned values whenever possible. To this end, Annex A specifies "where appropriate, the uncertainty of assigned values should be determined using procedures described in the *Guide to the Expression of Uncertainty in Measurement*[6]." This ISO/IEC Guide commonly called the "GUM" is the mother of all uncertainty publications. This guide helps ensure test artifact assigned value uncertainties are derived and reported using standardized formats so as to avoid confusion and computation errors.

The section of Annex A on performance statistics is divided into two main sections, performance on a single test item and performance on more than one test item for a particular measurand. Performance statistics are routinely used in interpreting proficiency test results. Proficiency test data transformed into performance statistics is easily evaluated in terms of deviation from target values (performance criteria). Some of the most common performance statistics computed from proficiency test data are derived from test data variability such as standard deviation, coefficient of variation, relative standard deviation, percentiles, etc. Other commonly used proficiency test performance statistics are derived from the difference between the test data and a test artifact's assigned value. This difference, often referred to as the "estimate of laboratory bias," is the most easily understood statistic by participants. The three most widely used proficiency test performance statistics as derived from the estimate of laboratory bias are percent difference, z scores and En numbers.

The percent difference performance statistic is simply the difference between a participant's test data and the test artifact's assigned value, divided by the test artifact's assigned value, multiplied by 100.

- **Percent difference:** $[(x - X) / X] * 100$ where x = participants test value, X = test artifact assigned value The z-score performance statistic takes the difference between a participant's test data and the test artifact's assigned value divided by a variability performance statistic such as standard deviation.

- **z-score:** $(x - X) / s$ where s = measure of variability

Note: The variability performance statistic used in the z-score computations should be based on enough observation as to reduce the influence of extreme test results.

The $E_n$ number (expressed as the "E sub N" number) performance statistic is derived by dividing the difference between a participant's test data and the test artifact's assigned value by the square root of the sum of the squares (RSS) of the participant laboratory's test data uncertainty and the reference laboratory's test artifact's assigned value uncertainty.

- **$E_n$: $(x - X) / (U2_{lab} + U2_{ref})$** where $U2_{lab}$ = participant laboratory test data uncertainty
  $U2_{ref}$ = reference laboratory test artifact assigned value uncertainty

The percent difference and the En number are normally computed and reported by a proficiency test provider upon receiving a participant's test data while z-scores are usually given in the final proficiency test report.

Proficiency tests based on more than one test item for a particular measurand normally involve test artifacts that are multifaceted and/or involve more than one test artifact. These types of proficiency tests usually report results graphically as an effective means for interpreting performance.

Two commonly used graphics techniques are Youden Plots and plots computed from Mandel's h-statistics. Other performance statistics used for proficiency tests based on more than one test item for a particular measurand are derived either from composite results from the same measurand or composite results from different measurands.

Evaluation of proficiency test performance is a critical part of each test's review process. Performance evaluation establishes whether or not the proficiency test objectives have been met by each of the participant laboratories. One of the first steps in the performance evaluation is to determine if proficiency test results are fit for the purpose for which they were intended. This activity, normally accomplished with the help of the proficiency test provider's technical advisors, reviews participants' method performance specifications, participants' recognized level of operations, and test artifact performance (stability and/or homogeneous considerations). After verifying that proficiency test results are fit for their purpose, performance statistics are evaluated to determine performance levels.

$E_n$ numbers and z score performance levels are normally determined as follows:

- **$E_n$:**
  - $E_n \leq 1$ = satisfactory performance
  - $E_n > 1$ = unsatisfactory performance
- **z-score:**
  - $|z| \leq 2$ = satisfactory performance
  - $2 < |z| < |z|$ = questionable performance
  - $|z| \geq 3$ = unsatisfactory performance

The consensus of participants may also be calculated, which is a central percentage at the 80%, 90%, or 95% level, in order to be satisfactory. Annex A recommends "graphs should be used whenever possible to show performance (e.g. histograms, error bar charts, ordered z-scores charts)."

The purpose of using charts and graphs is to illustrate the following:

1) Distribution of participant test data
2) Test data relationships between multiple test items
3) Distributions for different test methodologies

Another popular graphic technique, which gives a good visual representation of a participant's test data and associated uncertainty compared to a test artifact's assigned value and its associated uncertainty, is accomplished by displaying the boundaries created by participant's and test artifact's uncertainties. Furthermore, it evaluates whether there is overlap between them, and if overlap is present, to what degree (see Fig. 1.0).

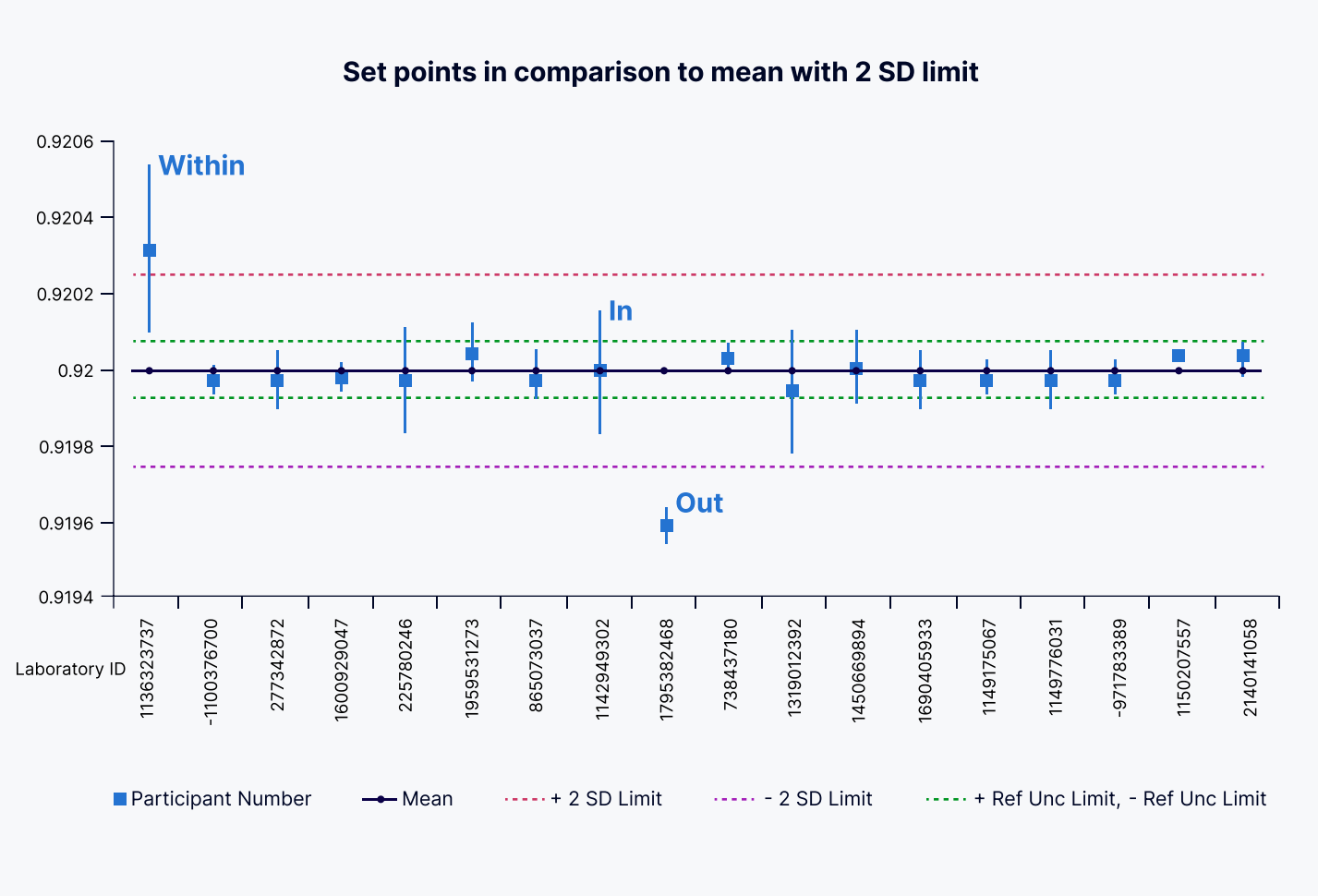
The following is commonly used to evaluate performance levels for this type of "uncertainty overlap" graph:

**I – In** This status is achieved when the participant's reported value falls ***within the uncertainty limits of the established reference value.*** The uncertainty of the participant has no bearing on this evaluation.

**W – Within** This status is achieved when the participant's reported value falls ***outside the uncertainty limits of the established reference value, but the uncertainty of the participant overlaps*** the uncertainty of the established reference value.

**O – Out** This status is achieved when the participant's reported value and reported uncertainty falls ***outside the uncertainty limits of the reference value.*** No agreement.

## UNCERTAINTY OVERLAP PLOT




Proficiency tests can, by design, include statistical techniques to monitor a participant's performance over time. These statistics can be used to determine a participant's performance variability, identify general trends, and spot inconsistencies. Laboratories often develop traditional "Shewhart" control charts from these statistics, which help facilitate their easy interpretation and allow improvements and problems to be readily identified.

## PROFICIENCY TESTING: LOOKING FORWARD

From an auditor's perspective, every laboratory should be taking time to perform evaluations to discover if documented evidence of demonstrated compliance is readily available. **Documented evidence of demonstrated compliance** refers to evidence that is in accordance with current and recognized quality standards. This process can prove somewhat taxing, but always worthwhile and beneficial to the security and growth of a laboratory. As the demand for accredited laboratory services increase, so will the demand to show competency in proficiency testing. This will transpire not only in accredited labs, but also from organizations that state they comply with recognized quality standards, i.e. ISO/IEC 17025.

Proficiency testing is a viable monitoring scheme which meets the requirements of ISO/IEC 17025 section 5.9 and is the most prolific means by which accreditation bodies establish and monitor a laboratory's demonstrated competence. Proficiency testing differs from traditional interlaboratory comparisons on several issues and involves considerable effort on behalf of a proficiency test provider to ensure test results are accurate, unbiased, fair, confidential, and most of all, useful to participant laboratories. Proficiency testing is indeed a powerful tool when applied to the workings of any laboratory seeking to stand behind all that they do.

## WORKS CITED


[1] ISO/IEC 17025. *General Requirements for the Competence of Testing and Calibration Laboratories*. 1999 Dec 15.

[2] NCSL Recommended Practice. *Guide for Interlaboratory Comparisons*. 1999 Mar, RP-12.

[3] ISO/IEC Guide 43-1 *Proficiency Testing by Interlaboratory Comparisons – Part: 1, Developing and Operation of Proficiency Testing Schemes*. 1997.

[4] ILAC *Guidelines for the Requirements for the Competence of Providers Testing Schemes*. 1999 Jul, Voting Draft 1.

[5] ASTM E 1301-95 *Standard Guide for Proficiency Testing by Interlaboratory Comparisons*.

[6] ISO *Guide to the Expression of Uncertainty in Measurement*. 1993.