# A DISCUSSION OF PROFICIENCY TESTING BENFITS AND PROGRAM STRATEGIES

Ken Middlebrook
Proficiency Testing Manager
Canadian Association for Environmental Analytical Laboratories (CAEAL)

## 1.0 INTRODUCTION

It is generally accepted that proficiency testing (PT) is a valuable quality assurance tool for testing laboratories. The standards that govern both testing laboratories (ISO/IEC 17025), and the accreditation bodies that accredit the testing laboratories (ISO/IEC 17011), make statements about participation in PT.

> ISO/IEC 17025 5.9.1 *The laboratory shall have quality control procedures for monitoring the validity of tests and calibrations undertaken…. This monitoring…may include…*
>
> > *b) participation in interlaboratory comparison or proficiency-testing programmes;*

> ISO/IEC 17011 7.15.3 *The accreditation body shall ensure that its accredited laboratories participate in proficiency testing…where available and appropriate, and that corrective actions be carried out when necessary.*

In spite of this, there are very few studies that have attempted to objectively measure the benefits of accreditation and proficiency testing to the testing laboratory and its clients. The main reason for this is the scarcity of objective, non-biased tools that can be used to measure these benefits.

The CAEAL PT program, and other similar programs, provide such a tool. This program has been operating since 1991, shipping PT samples to laboratories twice each year. Each study consists of four samples and the evaluation uses z scores estimated from consensus data. One of the strengths of the CAEAL program, with respect to using it for long-term trend examinations, is the way that data is collected and maintained. When a laboratory registers for PT, a unique PT code is assigned to each analyte; whenever a laboratory implements a new method for an analyte, a new PT code is assigned. As well, a laboratory registered for PT will continue to receive samples twice per year until they instruct otherwise. These administrative features result in a robust and traceable database.

## 2.0 Benefits of Accreditation

It is well understood that there are many different types of testing laboratories, with different clients and different quality expectations. Participants in CAEAL PT programs can be divided into three major categories,

1. Testing performed in water and wastewater treatment facilities that are not amenable to accreditation. This testing is performed in support of process control and is usually performed by operators with a limited knowledge of analytical chemistry, using pre-packaged test kits. As well, this testing is often conducted under conditions that are not normally considered acceptable for testing activities (e.g., poor ventilation, dust, vibrations, etc.).
2. Fully equipped testing laboratories with full-time technical staff that have opted not to pursue accreditation.
3. Accredited Laboratories
    a. Fully equipped laboratories that have obtained accreditation due to regulatory or client requirements but who have not fully embraced the principles. In short, these laboratories look on accreditation as a "necessary evil" and, although meeting the requirements of ISO/IEC 17025, do not fully understand the importance of these requirements.
    b. Fully equipped laboratories that have obtained accreditation, for whatever reason, but have embraced the principles of the standard and approach the process in a proactive and positive manner.

Intuitively, we would expect that the average quality of data produced in these laboratories would increase from scenarios one through three b, however, there are very few studies that have attempted to examine this.

### 2.1 ***Process Testing versus Analytical Testing (Scenario 1 vs 2)***

In addition to the regular PT program that CAEAL offers in support of dedicated testing laboratories, CAEAL also operates a PT program in partnership with Alberta Environment that is restricted to testing performed by operators in water and wastewater treatment facilities (process Laboratories). This data provides an excellent opportunity to compare the quality of process control testing against standard laboratory testing.

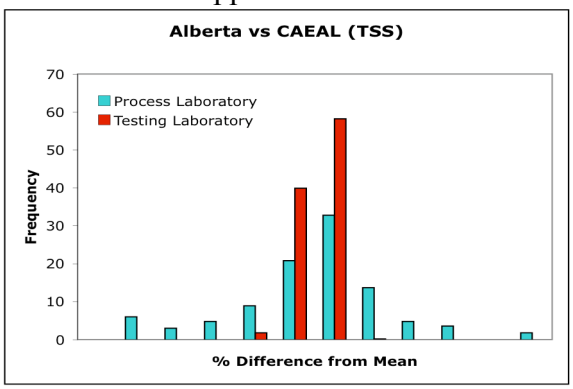


***Figure 1: Distribution of PT results for process control laboratories and for testing laboratories (total suspended solids).***

As expected, the distribution of results for an analyte in the Alberta program is typically wider than that observed in CAEAL's main program for the same

analyte (Figure 1). The one exception to this (residual chlorine) will be discussed later.

## 2.2 *Accredited versus Non-accredited Laboratories (Scenario 2 vs 3)*

Both scenario two (non-accredited laboratories) and scenario 3 (accredited laboratories), contain laboratories with equipment and facilities ranging from very small laboratories offering a limited number of tests to large, full service laboratories. Comparing performance in the CAEAL PT program between accredited and non-accredited laboratories clearly demonstrates that accredited laboratories perform better than non-accredited laboratories (Morris and Macey, 2004). This examination has been repeated several times with consistent results (Figure 2).

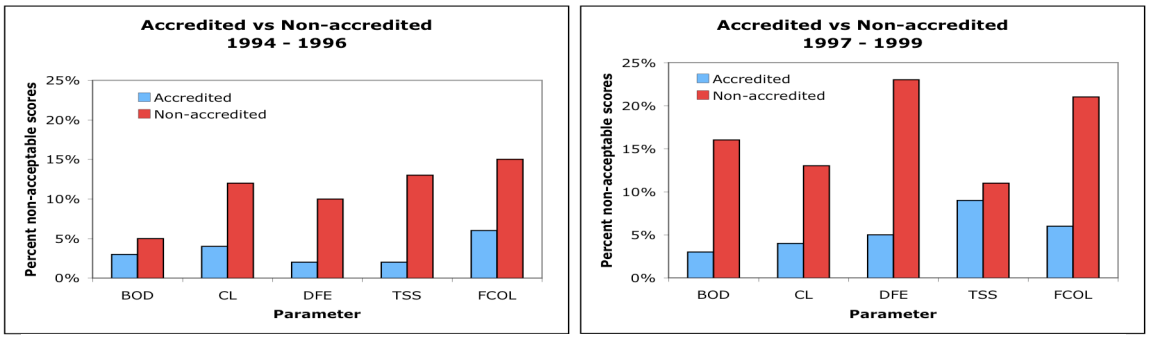


***Figure 2: Comparison of non-acceptable PT scores between accredited laboratories and laboratories that only participate in PT. BOD, CL, DFE, TSS and FCOL refer to biochemical oxygen demand, chloride, iron, suspended solids and fecal coliforms, respectively.***

Given the higher level of scrutiny on accredited laboratories, this observation is not unexpected.

## 2.3 *Differences Between Accredited Laboratories (Scenario 3a vs 3b)*

Accreditation is granted to any laboratory that has demonstrated conformance to ISO/IEC 17025. Although this process assures that all accredited laboratories meet a specified level of competence, it does not mean that all accredited laboratories are "equal". The quality of data will also be affected by the extent to which laboratory management embrace the concepts and principles of the standard, the degree of "buy-in" by laboratory staff, and the extent to which a laboratory goes above and beyond the requirements of the standard. Although this makes intuitive sense, it is very difficult to demonstrate empirically.

The CAEAL programs afford a unique opportunity to examine this assumption. CAEAL maintains a pool of volunteer assessors selected from participant laboratories. It can be argued that laboratories that make assessors available to CAEAL are, on average, more likely to be highly supportive of the accreditation process within their own laboratories than laboratories that do not. The rationale

for this argument is the very real costs incurred by laboratories that provide assessors. Each assessor may be away from the laboratory for up to two weeks per year; expenses are covered by CAEAL but salary costs are still covered by the laboratory.

A plot of the percentage of non-acceptable PT scores for accredited laboratories with active assessors (1,924 PT scores) and accredited laboratories without active assessors (2,845 PT scores) shows that, in general, laboratories with active assessors perform better than laboratories that do not provide assessors (Figure 3). Unfortunately, due to the means in which assessor information is recorded in the CAEAL database, it is not possible to extend this examination beyond the most recent PT study.

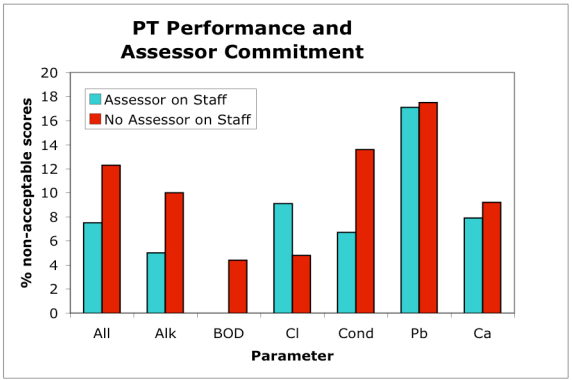


***Figure 3: Percentage of non-acceptable PT scores for accredited laboratories that provide assessors and accredited laboratories that do not (March 2005 study).***

## 3.0 Benefits of Proficiency Testing

The above examinations clearly show that PT data can be used to demonstrate the value of laboratory accreditation but we still have not shown that participation in a PT program, regardless of accreditation status, provides a quality advantage to laboratories.

In an attempt to address this issue, the entire set of PT data collected between March 1995 and March 2003 was examined. The resulting data set consisted of 65,161 separate PT scores over seventeen separate PT studies. All lab/analyte/method combinations that participated in CAEAL PT prior to the March 1995 study were removed, thereby ensuring that the first participation for any analyte was captured. Of the remaining data, only lab/analyte/method combinations that participated in at least ten consecutive studies were kept. The resulting data set consisted of 1,210 lab/analyte/method combinations and a total of 12,100 PT results.

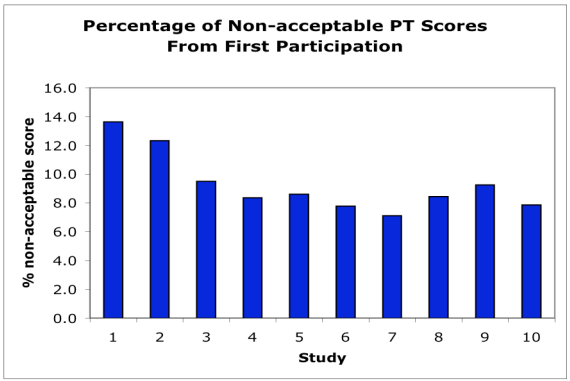


***Figure 4: Percentage of non-acceptable PT scores following participation in the first PT***

The remaining data was aligned such that the first participation was assigned to the first column in the array, the second participation to the second column and so on. The plot of percentage of non-

acceptable PT scores versus study shows that the performance improves over the course of the first several studies. (Figure 4).

A similar observation was made by David Hassemer of the Wisonsin State Laboratory of Hygiene (on-line report, 1996). Mr. Hassemer charted PT performance for several clinical analytes over eleven studies. The first three studies were conducted prior to the mandatory requirement for PT. The next study, the first that mandated PT for clinical laboratories, saw a significant increase in the percentage of non-acceptable PT scores, largely due to the addition of new participants.

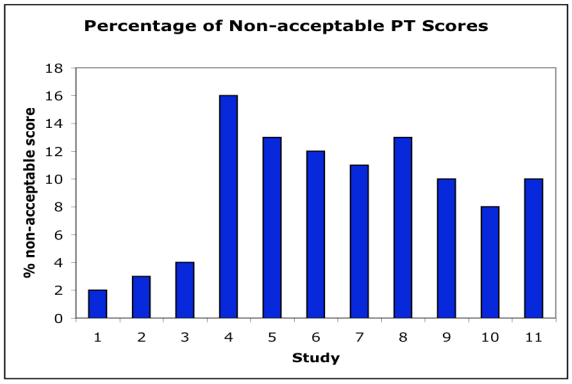


***Figure 5: Percentage of non-acceptable PT scores over several studies for cholesterol (Adapted from Hassemer 1996).***

Over the next several studies there was a consistent improvement in average PT performance (see Figure 5).

Both of these examinations clearly show that, on average, a laboratory's performance improves as the result of PT participation. Whether the improvement is the direct result of laboratories reviewing PT reports and acting on them or simply due to the increased awareness of quality issues within laboratories that goes hand in hand with participation in PT programs, is not certain.

## 4.0 FREQUENCY OF PT STUDIES

One of the most controversial topics related to proficiency testing is the frequency of studies. ILAC-P9:2005 provides guidance to accreditation bodies on the minimum frequency of PT. In short, this is one activity prior to granting accreditation and one activity relating to each major sub discipline within four years. A review of PT frequencies used by different bodies shows that they vary from the minimum suggested by ILAC to as many as ten studies per year (Figure 6). This discrepancy in PT frequencies amongst similar bodies makes it difficult to justify claims of equivalence between economies.

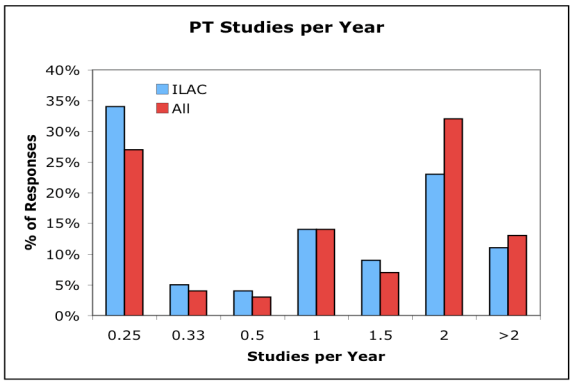


***Figure 6: Frequencies of PT studies from 56 ILAC accreditation bodies and 15 non-ILAC bodies.***

The PT frequencies employed by different accreditation and regulatory bodies are often based more on perceived risk to the public than on real, measurable benefit

to the quality of testing data. For example, PT requirements for environmental testing generally fall somewhere between annual and twice per year, whereas clinical and food testing PT is often conducted at frequencies of three to ten times per year. This in spite of the fact that,

> *"Only one reliable study on frequency has been reported [18], and that showed (in a particular scheme) that changing the round frequency from three to six per year had no significant effect (beneficial or otherwise) on the participants' performance." (The International Harmonized Protocol For The Proficiency Testing Of Analytical Chemistry Laboratories*; *Pure Appl. Chem.*, Vol. 78, No. 1, pp. 145–196, 2006),

The Harmonized Protocol goes further in stating that,

> *"If the period between distributions extends much beyond four months, there will be unacceptable delays in identifying and correcting analytical problems, and the impact of the scheme on the participants could be small. There is little practical value, in routine analytical work, in proficiency tests undertaken much less than twice a year."*

4.1 ***Why is Frequency Important?***

Proficiency testing is one of the key tools used to assess laboratory competence. Acceptable performance must be demonstrated prior to accreditation being granted and for ongoing maintenance of accreditation. The question that must be answered is, what level of surveillance is needed for an accreditation body to have confidence in granting and maintaining accreditation?

ISO/IEC 17011 states, *"The accreditation body shall ensure that its accredited laboratories participate in PT…and that corrective actions are carried out when necessary."* CAEAL accredited laboratories are required to submit a corrective action report for each unacceptable PT score. Failure to submit an acceptable corrective action report results in suspension of accreditation for the affected analyte. Even with this high level of scrutiny approximately twenty percent of unacceptable PT scores observed in a study had an unacceptable score in the previous study for the same analyte, indicating that the corrective action implemented after the previous unacceptable PT was not effective.

## 4.2 ***Risk Based PT Frequencies***

It is likely that the best approach to PT is one that takes into account the risk of producing suspect data. This should not be confused with risk to public health, although this must also be taken into account. A risk-based approach can be method specific or laboratory specific. A method specific system is based on the fact that inter-laboratory precision is often related to the method used. For example, testing that involves direct measurement using very simple equipment may be afforded a longer duration between studies than methods that are affected by numerous factors.

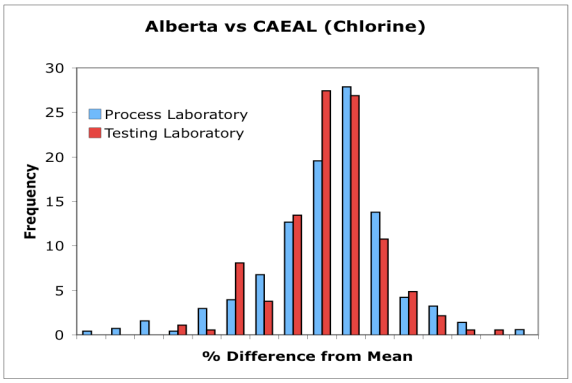


***Figure 7: Percentage of non-acceptable PT scores vs score in previous PT study (blue) and distribution of PT scores (red).***

This approach is supported by the observation made in the comparison between process laboratories and analytical laboratories. Although the precision of results observed for process laboratories are generally lower than that observed for analytical laboratories (see section 2.1), some analytes display very similar precisions (Figure 7). The reason for this is that some tests are inherently more robust and less dependent on analyst skill and knowledge than others.

Although the idea of risk based PT frequencies based on method is scientifically justified, the effort required to evaluate the robustness of all available methods may be prohibitive. As well, it does not take into account the very real differences that may be observed in some laboratories due to serious problems.

Another option would be a laboratory/analyte specific risk based approach. This approach is based on the fact that not all laboratories are equal and not all tests within a laboratory are equal.

Observation from historic CAEAL PT data show that laboratories that score high in a PT study are much less likely to obtain a non-acceptable score in the subsequent study (Figure 8). Reducing the frequency of PT for laboratories that consistently perform to a high level can be done without significantly increasing the risk. Because laboratory performance is analyte specific, this approach would have

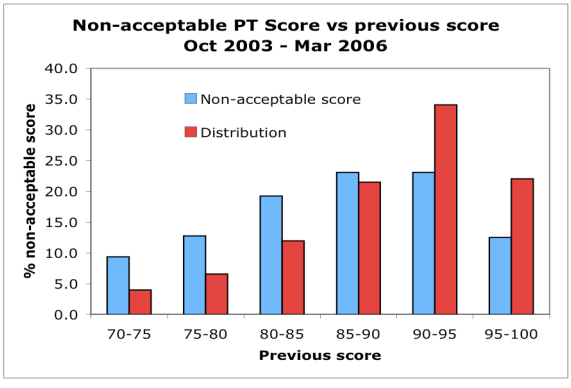


***Figure 8: Percentage of non-acceptable PT scores vs score in previous PT study (blue) and distribution of PT scores (red).***

to be managed on a lab/analyte basis.

## 5.0 Conclusions

This report shows that there are clear and measurable benefits to laboratory accreditation and to the participation in appropriate proficiency testing studies. Although the need to participate in proficiency testing is generally accepted by all accreditation bodies, there is no consensus on the optimal frequency. The current situation is that different accreditation bodies require different levels of PT participation. This creates an unequal and possibly unfair situation to bodies that are signatory to regional mutual recognition agreements.

Based on the above examination, several observations are made that are relevant to determining the most appropriate proficiency testing strategy,

- Not all laboratories, even amongst accredited laboratories, produce data of equal quality.
- Proficiency testing is an important surveillance tool for the granting and maintenance of accreditation.
- A significant proportion of corrective actions implemented as the result of an unacceptable PT score may not be effective.
- Laboratories that score very high (upper ten percent) in a PT study are disproportionately less likely to obtain an unacceptable score in a subsequent study than those that have a lower (acceptable) score.

The implication of these observations is that the appropriate frequency of PT is critical to assuring good quality laboratory data. The recommendation of a minimum PT frequency to cover all fields of testing, as provided in ILAC P09, is not sufficient. A risk-based approach to determining PT frequency is more scientifically justified. This could be a combination of establishing a minimum frequency that is field specific. For example, an appropriate frequency for environmental testing may be different than that desired for food or drinking water testing. These frequencies may be further modified on a laboratory/analyte specific basis by reducing the frequency for laboratories that consistently demonstrate a high level of performance.